PHD

The Pathogenesis of Clostridium difficile infection

Kirby, Jonathan

Award date:
2011

Awarding institution:
University of Bath

The Pathogenesis of *Clostridium difficile* infection

Jonathan M. Kirby

A thesis submitted for the degree of Doctor of Philosophy

University of Bath
Department of Biology and Biochemistry

May 2011

The more I know, the less I understand.
The more I understand, the less I know.
- Tao Te Ching

Table of Contents

I - List of Tables

II - List of Figures

III - Acknowledgements

First of all, I would like to thank my supervisors, Dr. Cliff Shone (HPA) and Prof. K. Ravi Acharya (University of Bath) whose freedom, support and guidance has been invaluable. While my path to this place may not have always been smooth, thank you for giving me the opportunity, I have loved it. I could've done without the mouldy house though!

I would like to thank, Dr. April Roberts and all those I have met in the Toxins team and around the HPA, I can't believe how fast the time has gone. I also offer my thanks to the Structural Molecular Biology group at the University of Bath with a special mention to Nethaji Thiyagarajan and Geoffrey Masuyer, thank you both.

Most of all, I would like to thank all those who have given me the opportunity to be in this valued position, many of whom will probably never read this. To many of these I also offer an apology, I am sorry if I have pushed you aside or not given you the time you deserve.

IV - Abstract

Clostridium difficile is a major problem as the aetiological agent of antibiotic associated diarrhoea. The mechanism by which the bacterium colonises the gut is poorly understood, but undoubtedly involves a myriad of components present on the bacterial surface. The aims of this study were to further define roles for selected surface proteins using a knockout approach, to evaluate the feasibility of surface protein based immunotherapeutics and to obtain structural information using X-ray crystallography. Mutants of cell wall-binding domain (PFam04122) containing proteins CD1036, CD2735, CD2784, Cwp66, CD2791, Cwp84, CD2795 and the flagella cap (FliD) were created. Mutants were characterised with regard to growth, sporulation, toxin production, adhesion *in vitro*, and, for the Cwp84 mutant, using the *in vivo* hamster model. The surface-located cysteine protease, Cwp84, was found to play a key role in maturation of the *C. difficile* S-layer, yet the Cwp84 mutant still caused disease with a similar pathology to the wildtype. Culture supernatant levels of toxin A were increased in CD2735, Cwp66, CD2791, CD2795 and particularly in Cwp84 and FliD 24 hr cultures, while CD2735, Cwp66, CD2791, CD2795 mutants also showed reduced adherence to Caco-2 cells compared to the wild-type. Passively administered immunotherapy, generated to low pH surface protein extracts of the *C. difficile* R20291 strain, did not protect hamsters from challenge with the cognate strain. Structural studies were undertaken on the surface proteins CD2791, Cwp66 and CD2767. Crystallisation conditions were identified for a recombinant N-terminal domain of CD2767 and an X-ray data set collected to 2 Å, although the structure was not solved by molecular replacement. Together these results further our knowledge of *C. difficile* surface proteins, although further work is required to identify which surface proteins play key roles *in vivo* during infection.

V- List of abbreviations

(k)Da	(kilo) Dalton
aa	Amino acid
ABP	Activity Based Probe
amp	Ampicillin
ARL	Anaerobe Reference laboratory
bis-tris	bis-(2-hydroxy-ethyl)- amino-tris(hydroxymethyl)-methane
BLAST	Basic Local Alignment Search Tool
bp	Base Pair
BSA	Bovine Serum Albumin
carb	Carbenicillin
CCD	Charged couple device
CDI	*C. difficile* infection
cDNA	Complementary DNA
CDT	*C. difficile* binary toxin
CDTa (*cdtA*)	Enzymatic component of *C. difficile* binary toxin
CDTb (*cdtB*)	Transport component of *C. difficile* binary toxin
CDΔgene	*Clostridium difficile* deficient in gene proceeding delta
cfu	Colony forming unit
Clos	ClosTron
CWBD	Cell Wall Binding Domain (PFam 04122)
Cwp	Cell Wall Protein
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate (aka nucleotides)
DPBS	Dulbecco's Phosphate Buffered Saline
DTT	Dithiothreitol
E64	(2S,3S)-3-(N-((S)-1-[N-(4-guanidinobutyl)carbamoyl]3-methylbutyl)carbamoyl)oxirane-2-carboxylic acid
ECM	Extracellular Matrix
EDTA	Ethylenediamine tetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
Erm	Erythromycin
FAA	Fastidious Anaerobe Agar
FITC	Fluorescein isothiocyanate
gDNA	Genomic DNA
FOM	Figure of Merit
GeLC-MS/MS	Gel-based Liquid Chromatography- tandem Mass Spectrometry/Mass Spectrometry
GnHCl	Guanidine Hydrochloride
GST	Glutathione S-transferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIC	Ion Exchange Chromatography

His6	Hexahistidine
HMW	High Molecular Weight
HRP	Horse Radish Peroxidase
IEX	Ion Exchange Chromatography
Ig	Immunoglobulin
IL	interleukin
IMAC	Immobilized Metal Ion Affinity Chromatography:
IPTG	Isopropyl β-D-thiogalactoside
IVIG	*In vitro* Immunoglobulin
IVT	*In vitro* translation
kan	Kanamycin
KLH	Keyhole Limpet Hemocyanin
KO	Knockout (aka Mutant)
LB	Luria Bertani media
LMW	Low Molecular Weight
MAT	Metal Affinity Tag (HNHRHKH)
MBP	Maltose binding protein
MES	2-(N-morpholino)ethanesulfonic acid
MIB	Sodium malonate, Imidazole, and Boric acid
MMT	Malic acid, MES and tris base
mRNA	Messenger RNA
MSCRAMs	Microbial Surface Components Recognizing Adhesive Matrix Molecules
MW	Molecular Weight
MWCO	Molecular weight cut off
NHS (resin)	N-hydroxysuccinimide
OD	Optical Density
ORF	Open Reading Frame
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate Buffered Saline
PBS/T	PBS plus 0.1% Tween 20
PCR	Polymerase Chain Reaction
PDB	Protein data bank
PEG	Poly Ethylene Glycol
pI	Isoelectric point
PMC	Pseudomembranous colitis
PVDF	Polyvinylidene fluoride
qPCR	Quantitative PCR (aka Real time PCR)
r[PROTEIN]	Recombinant form of protein e.g. rCwp84
RAM	Retrotransposition-Activated Marker
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
RT-PCR	Reverse Transcription PCR
sBHI	Supplemented Brain Heart Infusion
SCP	Sperm coat protein aka Serine chelating protease
SCWP	Secondary Cell Wall Polymer

SDS	Sodium dodecyl (lauryl) sulphate
SH3	SRC Homology 3 Domain
SLP	Surface layer Protein
SLPs	Surface Layer Proteins
SOE PCR	Splicing by overhang (overlap) extensions
SUMO	Small Ubiquitin-like MOdifier
TB	Terrific broth media
TcdA (*tcdA*)	*C. difficile* Toxin A
TcdB (*tcdB*)	*C. difficile* Toxin B
TIM	triosephosphateisomerase
TMB	Tetramethyl benzidine
TT	TargeTron
WT	Wild-type

1 General Introduction

Clostridium difficile is a Gram-positive, spore-forming rod-shaped bacterium, first identified in 1935 (Hall & O'Toole 1935) but is thought to have evolved within the last 1.1 - 85 million years (He *et al.* 2010). *C. difficile* can be cultured from the stool of 2-3% of healthy adults and up to two thirds of healthy newborns and infants (Viscidi *et al.* 1981; Larson *et al.* 1982; Sunenshine & McDonald 2006). *C. difficile* infection (CDI) is primarily associated with the administration of antibiotics, most famously clindamycin (Bartlett *et al.* 1977), but particularly broad spectrum penicillins, third generation cephalosporins and fluoroquinolones (Bartlett 2006). CDI symptoms range from mild self-limiting diarrhoea without appreciable lesions through colitis, pseudomembranous colitis (PMC) and finally, if left untreated, CDI culminates with colonic perforation, toxic megacolon, prolonged ileus, ascites and death (Kelly & LaMont 1998). CDI has a mortality rate typically between 6-30% (Bartlett 2002). Metronidazole or vancomycin (identified originally to treat *S. aureus* colitis (Bartlett 2008a)) are, in that order, the usual treatment regimens. However, between 10-40% of patients relapse (McFarland 2008), thus patients not only require repeated rounds of treatment, but often suffer repeated rounds of symptoms.

The unique environment found in healthcare settings (large cohorts of patients, wide antibiotic use, age of inhabitants and a transient 'population') provides a discrete 'ecosystem' where *C. difficile* (primarily spores) persist and can be rapidly transmitted amongst individuals. Increased stay in this environment is but one of a number of risk factors for CDI. These include, age, gender, use of proton pump inhibitors or immunosuppressant's and underlying disease e.g. inflammatory bowel disease (Vaishnavi 2009).

The numbers of patients suffering from CDI have increased over the past decade (increases in reporting and diagnosis aside) (Figure 1.1). In the United States, CDI is responsible for more deaths than all other intestinal infections combined (Redelings *et al.* 2007) causing considerable economic cost (Ghantoji *et al.* 2010).

The increase in incidence and mortality was particularly marked by an epidemic in 2003-2004 in Quebec, Canada (Pepin *et al.* 2005). This outbreak was characterised by CDI that was more severe, more refractory to therapy and subject to high rates of relapse (Bartlett 2006).

Upon investigation of isolates, it appeared most patients were infected with one particular, previously uncommon, fluoroquinolone resistant strain, *C. difficile* BI/NAP1/027 toxinotype III (restriction-endonuclease analysis [REA] group BI, pulsed-field gel electrophoresis [PFGE] type NAP1 [North American PFGE type 1] and polymerase chain reaction (PCR) ribotype 027]) (Pepin *et al.* 2005; MacCannell *et al.* 2006)

The 027 ribotype has since been isolated worldwide and now represents the most frequently isolated *C. difficile* ribotype in the UK (Brazier *et al.* 2008; Freeman *et al.* 2010; Gerding 2010). Based on this ability i.e. rapid spread, and particularly its association with more served disease, there remains concern about whether any other strain(s) e.g. ribotype 078 (Burns *et al.* 2010c) can cause another epidemic.

In the UK, deaths from *C. difficile* appear to be decreasing (Figure 1.1). Reasons behind this are likely to be increased awareness of CDI and increased diagnosis and treatment; however the effect of hospital cleaning and antibiotic stewardship should not be underestimated (Vonberg *et al.* 2008).

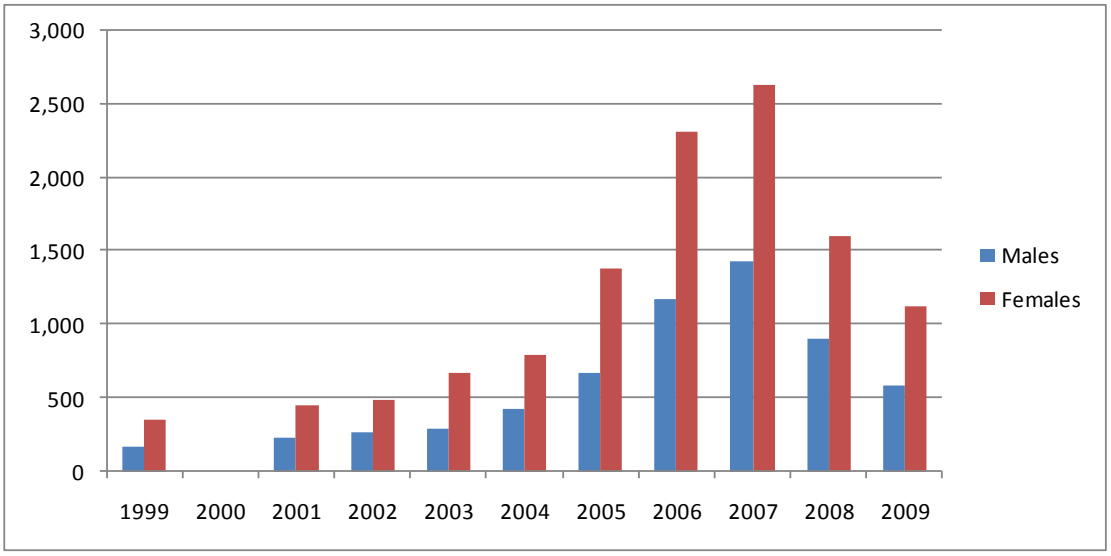


Figure 1.1 Numbers of deaths in the UK where *C. difficile* was the underlying cause of death. Source: Office for National Statistics (ONS) data.

1.1 Pathogenesis

C. difficile is primarily acquired via spores from a range of surfaces, the skin of healthcare workers or from aerosolized spores from symptomatic patients (McFarland 2002; Bobulsky *et al.* 2008; Best *et al.* 2010). Following traversal of the stomach, putative spore receptors come in contact with bile acid derivatives (cholate and chenodeoxycholate) and begin the germination process: a process absolutely requiring SleC a spore cortex-lytic enzyme (Burns *et al.* 2010b; Ramirez *et al.* 2010; Sorg & Sonenshein 2010).

The constituents of the spore have been determined proteomically and upon comparision with other Clostridial spore there appears to be a set of 28 proteins unique to *C. difficile* including a collagen-like glycoprotein and BclA (CD0332), two surface layer proteins (CD2791 and CD2793 [SlpA]), putatively involved in in spore attachment or spore-host interactions (Lawley *et al.* 2009b). Interestingly, *C. difficile* spores appear covered in spines/hair like projections (Kumar, V. & Roberts, A.K., unpublished data) putatively involved in adhering to host cells, as demonstrated using field emission scanning electron microscopes (FESEM) (Panessa-Warren *et al.* 2007).

The proteolytic load and transit time of the small intestine generally prevents *C. difficile* colonisation. Therefore, in humans, the slow moving large intestine is the most common site of CDI infection (Keel & Songer 2006).

Administration of broad spectrum antibiotics, altered gastrointestinal function e.g. inflammatory bowel disease or immunodeficiency, removes competitive inhibition of *C. difficile*; however *C. difficile* outgrowth only occurs when the causative antibiotic goes below the MIC of *C. difficile* (Freeman *et al.* 2007).

C. difficile vegetative cell adhesion and colonisation mechanisms are still relatively unknown. To date only a limited number of putative adhesins have been identified, including the S-layer, flagella and other specific proteins (Deneve *et al.* 2009b). Bacterial adherence is arguably the first step in pathogenesis, assisting in the efficient delivery of toxins to the cell, thereupon putatively increasing adherence. For example the *C. difficile* actin-ADP-ribosylating (binary) toxin putatively causes protrusions that may aid bacterial adhesion (Schwan *et al.* 2009).

The primary virulence factors of *C. difficile* are its toxins, TcdA (*tcdA*), TcdB (*tcdB*) (herein referred to as toxin A and B) and a binary toxin (made up of two, CdtA and B subunits). Both toxin A (308 kDa) and toxin B (269 kDa) are cytotoxic and encoded on the 19.6 kbp pathogenicity locus (PaLoc) (Hammond & Johnson 1995). The PaLoc also encodes three other genes tcdR (previously named *tcdD* (Rupnik *et al.* 2005)), *tcdE* and *tcdC*. TcdR is an alternative RNA polymerase sigma factor that positively regulates toxin gene transcription and its own transcription. TcdC is a negative regulator (Dupuy *et al.* 2008) and TcdE is a putative holin, which may aid release of *C. difficile* toxins to the extracellular environment (Tan *et al.* 2001)

Both toxins A and B have four major domains: the glucosyltransferase N-terminal domain, translocation (delivery) and cysteine protease automaturation domains and a C-terminal receptor binding domain (Pruitt *et al.* 2010). Toxin A and B enter the cell through receptor-mediated endocytosis and require an acidified endosome for translocation. Following internalisation, toxin A and B inactivate small GTP-

binding proteins, which include Rho, Rac, and Cdc42, through monoglucosylation of a single threonine (Thr37 for RhoA) within the effector-binding loop and coordinate a divalent cation, critical to binding GTP, thus locking them in an inactive confirmation (Voth & Ballard 2005). Since Rho, Rac, and Cdc42 are involved in signal transduction associated with apoptosis and the maintenance and regulation of the cell actin cytoskeleton, loss of this regulation at the cellular level results in shrinking and rounding of cells via retraction of neurite-like fibres and finally, cell death (Poxton *et al.* 2001; Voth & Ballard 2005). Cordially, this cytoskeletal perturbation and loss of cell polarity results in breakdown of tight junctions, resulting in increased paracellular permeability of mucosal surfaces. The gut mucosa damage and toxin induced inflammation results in CDI aetiology (Ng *et al.* 2010).

The binary toxin (actin-specific ADP-ribosyltransferase) consists of two subunits, CDTa- the enzymatic component and CDTb- the binding component. The structure of CDTa has been determined (Sundriyal *et al.* 2009). The role of the binary toxin in pathogenesis is not clearly understood as 'hypervirulent' 027 strains produce the binary toxin (Cartman *et al.* 2010), however strains which do not express the binary toxin still cause disease e.g. 630 and VPI10463 (Rupnik *et al.* 2003). Conversely a pathogenic A-B- binary toxin-positive strain of *C. difficile* strain has been reported (Elliott *et al.* 2009) contradicting data that A-B- binary toxin-positive strains, while being enterotoxic, do not cause disease in hamsters (Geric *et al.* 2006).

Current understanding, given the lack of natural A+B- isolates, is that toxin B is required for virulence (Lyras *et al.* 2009). However, a *C. difficile* 630 (low toxin producer) A+B- knockout demonstrates virulence in hamsters, although colonisation to death figures appear more protracted compared to A+B+ or A-B+ strains (Kuehne *et al.* 2010).

As demonstrated using mice, CDI results in a so called supershedder state where spores are, unsurprisingly, transmitted into and contaminate the environment (Lawley *et al.* 2009a). Studies have suggested epidemic strains show increased toxin or prolonged production and robust sporulation, however these virulence factors

have been found to vary amongst BI/NAP1/027 isolates, and differ little from other non-epidemic strains (Akerlund *et al.* 2008; Burns *et al.* 2010a; Merrigan *et al.* 2010).

1.2 Bacterial surface proteins

Adherence and subsequent colonisation represents a key milestone in infection as without adhesion, bacteria will simply be excreted. *C. difficile* endogenous in asymptomatic carriers are, for example, already utilising the tools of adherence but have, as yet, not caused CDI.

Bacterial adherence requires the participation of two factors: a receptor or a ligand and an adhesin. Adhesins are typically found on the bacterial surface and can interact in a complementary and/or specific manner with their receptor or substrate.

The surface of Gram-Positive bacteria contains a plethora of proteins attached to the cell envelope in a variety of different ways (Desvaux *et al.* 2006) (Figure 1.2).

These include:

- Membrane proteins- associated with the phospholipid bi-layer
- Lipoproteins- containing lipid covalently linked to an N-terminal cysteine residue (Hutchings *et al.* 2009)
- LPXTG containing proteins –covalently linked to the peptidoglycan by sortases (Marraffini *et al.* 2006)
- Other putatively non-covalently attached surface proteins - which include domains putatively attaching/associating them to peptidoglycan e.g. LysM Motif (Buist *et al.* 2008), S-layer homology (SLH) domain, Cell Wall Binding Domain Type 2 (CWBD2) (PFam 04122)
- Motility associated proteins - flagella and pili (Jarrell & McBride 2008; Kline *et al.* 2010)

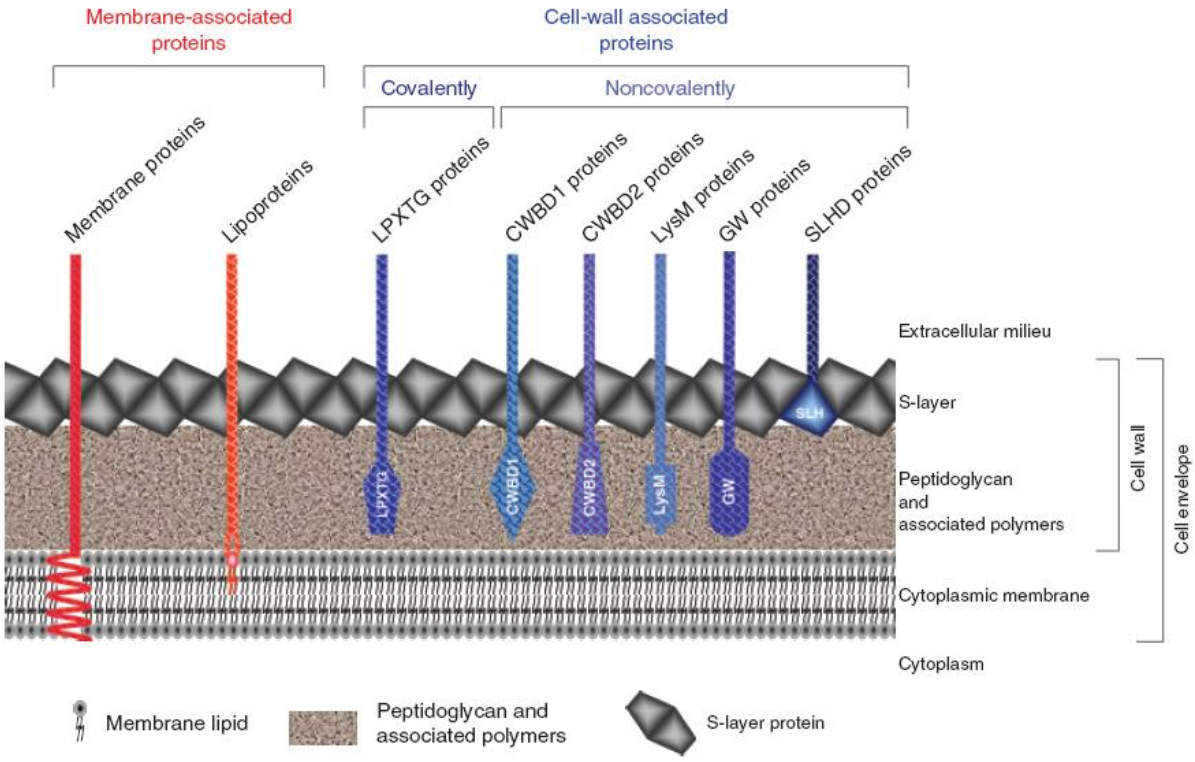


Figure 1.2 Schematic representations of the different types of surface proteins found in Gram-positive bacteria. Figure taken from Desvaux *et al.* (2006).

Interestingly, the Clostridial and Bacilli peptidoglycan is suggested to be thinner than other Gram positive species e.g. Corynebacterium, Mycobacterium, or Propionibacterium, and has been found to thin beneath the S-layer during lag and initial exponential growth phases. By stationary phase, Clostridia and Bacilli cultures can be virtually Gram negative (Beveridge 1990).

1.2.1 S-layers

The first and largest (in terms of cell coverage) of these surface proteins in *C. difficile* is that of the paracrystalline S-layer. S-layers are two-dimensional (glyco)protein lattices formed by the entropy driven aggregation of monomer subunits (Sara & Sleytr 2000). Despite the apparent local entropy decrease upon forming a regular array, the net entropy of the system is increased due to the release of water molecules from a hydration shell formed on the surface of the hydrophobic domains of the S-layer monomeric units (Teixeira *et al.* 2010). S-layers are found in a variety of pathogenic bacteria including *Bacillus anthracis* (the causative agent of Anthrax), *Campylobacter sp., Aeromonas salmonicida* and many

Archaea (Sara & Sleytr 2000). Maintenance of the S-layer over the entire bacterial surface requires significant energy investment, for example at a generation time of 20 minutes, approximately 500 S-layer subunits per second are needed to completely cover the bacterial cell (Sleytr & Messner 1983; Messner & Sleytr 1992). S-layers therefore provide a vital function to the bacterial cell for such an investment to be worthwhile. Engelhardt (2007) proposes S-layers provide basic primary 'survival' functions such as mechanical, osmotic and thermal stability while also providing functions such as protection against the immune system, adhesion and nutrient acquisition. Some S-layers contain Surface-Layer Homology (SLH) domains which are proposed to attach the S-layer subunits to the cell surface (Lupas 1996), although the method of actual attachment varies among species and may involve secondary cell wall polymers (SCWPs) (Mesnage *et al.* 2000; Schaffer & Messner 2005).

1.2.2 *C. difficile* S-layer structural identification and characterisation

The *C. difficile* S-layer was initially characterised by a group at Tokushima University School of Medicine, Japan. Analysis of autolysed cell walls revealed a regularly arranged rhombic array possessing sides of 8.1 nm and an interior angle of 88 degrees visible on the cell surface. SDS-PAGE analysis of cells treated with 4 M guanidine hydrochloride or 6 M urea revealed two main proteins of approximately 45 kDa (high molecular weight S-layer protein, HMW SLP) and 32 kDa (low molecular weight S-layer protein, LMW SLP) that varied in size amongst strains. Cell wall extracts were also able to form flattened or rod-like paracrystalline sheet-like fragments *in vitro,* a feature that appeared to require divalent cations specifically Ca^{2+} or Zn^{2+}. Resistance of the 35 kDa protein to trypsin degradation lead to the suggestion that, together with the 45 kDa proteins, the array formed by these two proteins played a role in protection against physicochemical agents including proteolytic enzymes. Cerquetti *et al.* (2000), using freeze etching of whole bacterial cells, demonstrated the presence of a square ordered lattice on the outermost surface and, using autolysed cell wall and urea extracted SLPs, an inner hexagonal lattice innermost with spacing of ~10 nm.

Moreover, antibodies to the either the LMW or HMW SLP, or *C. difficile* cells which had been subjected to treatment with urea or guanidine hydrochloride had significantly decreased adhesion *in vitro*, suggesting the S-layer played a key role in adhesion (Takumi *et al.* 1987; Takeoka *et al.* 1991; Takumi *et al.* 1991; Takumi *et al.* 1992). This was later supported by Karjalainen *et al.* (2001) who also found that antiserum to the LMW SLP inhibited adhesion of *C. difficile* to Caco-2 cells.

Other groups suggested that the size variability of the S-layer bands (also extractable with EDTA) formed a typing system (Wexler *et al.* 1984; Delmee *et al.* 1986; Delmee *et al.* 1990; McCoubrey & Poxton 2001). The typing of bacterial strains based on their S-layer (although the gene rather than the surface extracts) is still suggested today (McCoubrey & Poxton 2001; Karjalainen *et al.* 2002; Kato *et al.* 2005; Kato *et al.* 2009).

In 2001, as a result of searching the unfinished *C. difficile* 630 genome (completed in 2006 (Sebaihia *et al.* 2006)), it was found that the HMW and LMW SLPs were encoded on one gene (*slpA*) and post-translationally cleaved to their respective SLPs, a feature unique to *C. difficile*. The HMW SLP was found to contain repeat regions with homology to N-acetylmuramoyl-L-alanine amidase CwlB of *Bacillus subtilis* and demonstrated this activity in zymogram assay. The amidase domain was found in 28 other proteins at either the N- or C- terminus with a 'functional' domain at the opposite end, some of which have been found to be expressed during normal growth (Calabi *et al.* 2001; Karjalainen *et al.* 2001). The cell wall binding 2 domain (Pfam 04122) (CWBD) is thought to bind the overall protein to the cell surface, non-covalently, analogous to the S-layer Homology Domain (SLH (Lupas *et al.* 1994)) found in many species paracrystalline S-layers (Engelhardt & Peters 1998). The PFam 04122 domain appears not to be restricted to *C. difficile,* as 20 proteins containing the PFam 04122 are found in *Finegoldia magna* (formerly *Peptostreptococcus magnus*) (Goto *et al.* 2008) and many other bacterial species.

Patients appear to raise antibodies to both the HMW and LMW SLP during CDI (Pantosti *et al.* 1989; Cerquetti *et al.* 1992; Drudy *et al.* 2004; Pechine *et al.* 2005a; Wright *et al.* 2008). Responses to the HMW SLP are conserved across strains, while responses to the LMW SLP are only similar within a serogroup (Cerquetti *et al.* 2000; Calabi *et al.* 2001; Karjalainen *et al.* 2002). This variability is reflected genetically as, other than a conserved cleavage site between the two SLPs, the HMW portion of *slpA* has been found to be more conserved than the LMW (Eidhin *et al.* 2006).

Proteomic analysis of a range of different surface protein extraction techniques by Wright et al (2005), confirmed suggestions by Wexler *et al.* (1984) that in addition to the S-layer a large number of other proteins are extractable from the cell surface, including a number of the *slpA* paralogs. Removal efficiency of the HMW vs LMW SLP also differed, confirming the hypothesis that the LMW SLP was outermost, while the HMW SLP was intimately associated with the underlying cell wall.

Various studies had suggested that either one or both of the S-layer subunits were glycosylated (Mauri *et al.* 1999; Cerquetti *et al.* 2000; Calabi *et al.* 2001). However, a study by Qazi *et al.* (2009) utilised a range of methods to determine that the observed molecular masses matched the predicted masses of the LMW and HMW SLPs in a range of C. difficile strains, suggesting no glycosylation or posttranslational modifications other than cleavage of SlpA into the LMW and HMW SLP.

In 2009, a crystal structure of a fragment of the LMW SLP to 2.4 Å (PDB 3CVZ) was reported, together with data further characterising the interactions between the HMW and LMW SLP (Fagan *et al.* 2009). The crystal structure of the LMW SLP fragment revealed a two domain structure (Figure 1.3A), domain 2 (pink) displays a novel protein fold not yet classified (which is surface-most) while domain 1 (yellow) adopts a two-layer sandwich architecture.

By mixing recombinant HMW and LMW SLPs in an equimolar ratio, it was demonstrated that, similar to low pH extracted SLPs (Calabi *et al.* 2001); they form a single species (termed the H/L complex) in native PAGE. Small angle X-ray scattering (SAXS) in solution, using recombinant SLP subunits and extracted SLPs, revealed that the H/L complex was made up of an elongated heterodimeric molecule with approximate dimensions of 240 Å x 105 Å x 75 Å, with an estimated MW of 81 kDa, where the HMW and LMW interact 'end-to-end' (via a contact area of c. 450 Å²) presenting the LMW externally (Figure 1.3B). However, when low pH extracted SLPs were run on size exclusion chromatography (SEC) a single peak was observed corresponding to 320 kDa, thought to be caused by the elongated shape of the H/L complex. It was also demonstrated that the C-terminal 62 residues (260–321) of the LMW SLP were absolutely required to form the H/L complex, confirming the existence of a conserved cleavage site and SLP subunit interaction domain (Calabi & Fairweather 2002; Eidhin *et al.* 2006). Together the data suggests that the formation of the each H/L complex does not require co-factors and does not preclude that divalent cations or a currently unknown co-factor or cell wall component may be involved in S-layer array assembly, i.e. assembly of multiple H/L complexes or the non-covalent interaction with the cell wall.

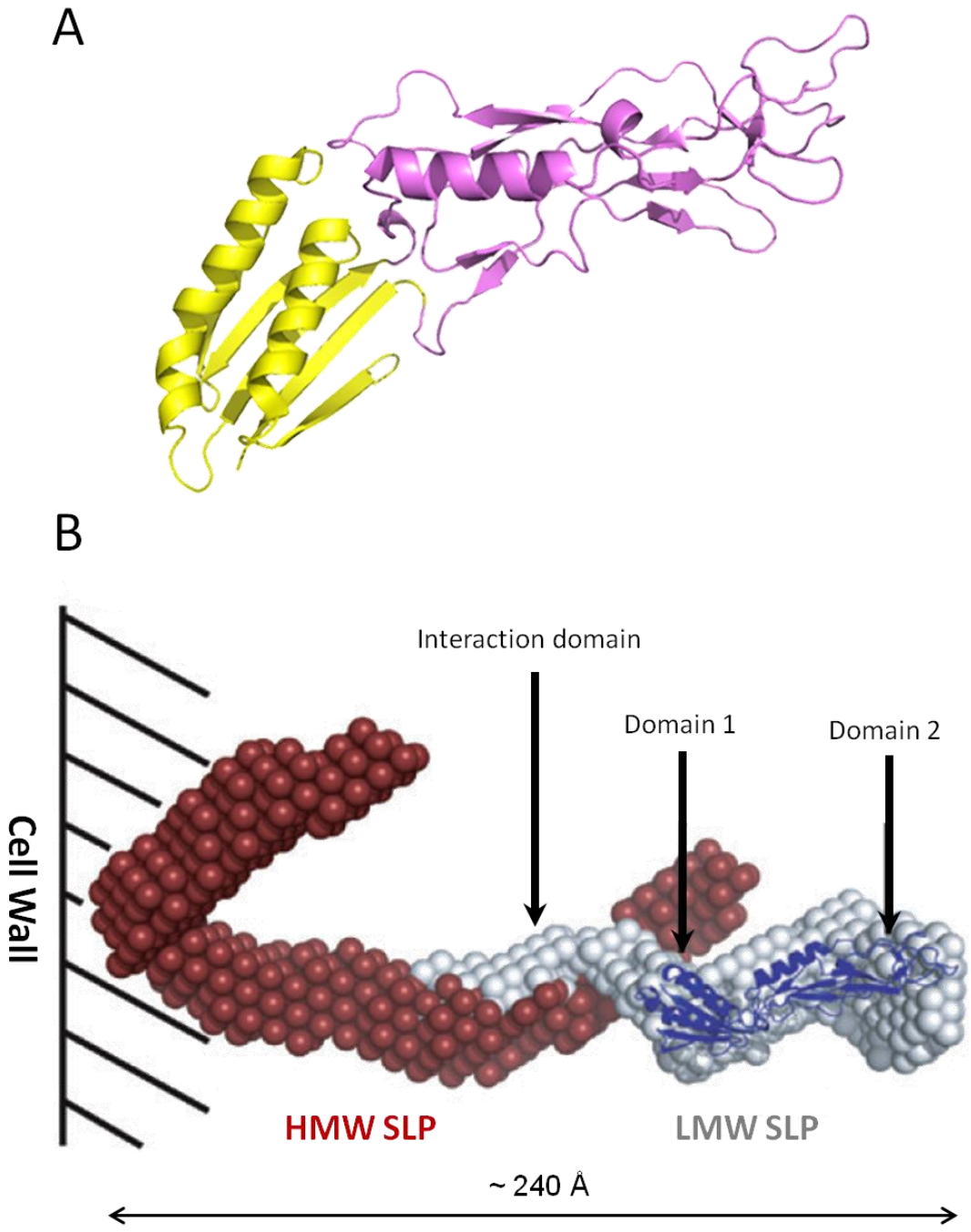


Figure 1.3 Proposed structures of the *C. difficile* S-layer. (A) Crystal structure of the N-terminal fragment of the LMW SLP (LMW_{1-262}) Domain 1 – Yellow, Domain 2 – Pink (B) Proposed orientation of the S-layer of *C. difficile,* combined LMW/HMW SAXS structure with overlaid LMW fragment crystal structure (Blue) (Note the crystal structure lacks the C-terminal region (residues 245–274 and 304–321) of the LMW SLP essential for LMW/HMW interaction for H/L complex formation. (A) Prepared using PyMOL (www.pymol.org) (B) Adapted from Fagan *et al.* (2009).

1.2.3 Role of the *C. difficile* S-layer and its proteins in infection

C. difficile cells, from which the S-layer had been removed by urea or GnHCl, scarcely adhered to human HeLa cells. Furthermore, adhesion could be inhibited by ~50% using FAb_2 fragments of anti-LMW or anti-HMW SLP IgG, however when used together they inhibited adhesion by 84% (Takumi 1991). Together these results suggested that the S-layer (and other removable surface proteins) play a key role in adhesion of *C. difficile*.

Recombinant HMW SLP and low pH extracted SLPs bound to gastrointestinal tissues, collagen I, thrombospondin and weakly to vitronectin, suggesting S-layer adhesion is ECM based, mediated by the HMW SLP (Calabi *et al.* 2002). Moreover, *C. difficile* SLPs appear to induce cytokines in monocytes and dendritic cells, suggesting that SLPs are able to affect inflammatory and regulatory cytokine production (Ausiello *et al.* 2006).

Together these results suggest that the *C. difficile* S-layer provides not only physiochemical stability to the cell but also mediates bacterial cell adhesion. It appears that an incorrectly matured S-layer i.e. no longer cleaved into the HMW/LMW SLPs by Cwp84, does not affect the ability of *C. difficile* to cause infection in the hamster model (Kirby *et al.* 2009), bringing into question the role of the S-layer in pathogenesis. Indeed, Calabi *et al.*, (2002) suggests that a *slpA* mutant would confirm the role of the S-layer in pathogenesis, although to date this mutant has proved elusive.

S-layer mutants of *Campylobacter fetus* subsp. Fetus (created by multiple passage (Tummuru & Blaser 1992)) were avirulent in a mouse model (Blaser & Pei 1993) and did not colonise or cause abortions in pregnant ewes when administered subcutaneously (Grogono-Thomas *et al.* 2000). BslA mutants of *B. anthracis* Ames (constructed by allelic exchange) were largely avirulent in subcutaneously challenged guinea pigs, despite toxin secretion remaining unchanged (Kern & Schneewind 2009). Furthermore, an S-layer (A-layer) mutant of *Aeromonas*

salmonicida was obtainable after growth of cultures at an increased temperature (Ishiguro *et al.* 1981) and were found to be more sensitive to bactericidal activity of serum (Munn *et al.* 1982) and avirulent in animal studies (Kay & Trust 1991). These studies therefore suggest that if a *C. difficile* SlpA mutant was created, it is likely to be less virulent than the WT.

Proteins on the cell surface of *C. difficile* putatively change in response to environmental stimuli e.g. heat shock thereby increasing adherence to cell lines (Eveillard *et al.* 1993; Karjalainen *et al.* 1994; Waligora *et al.* 1999). The *C. difficile* response to heat shock led to the characterisation of the 66 kDa cell surface *slpA* paralog Cwp66 (Waligora *et al.* 2001) (discussed further in Chapter 5).

Furthermore, the 160 kDa phase variable surface protein CwpV is, similar to SlpA, post-translationally cleaved into two interacting subunits, a 40 kDa (cell wall anchoring) and a 120 kDa (surface exposed) subunit. Mutants in which CwpV is permanently 'on', have altered colony phenotypes and increased bacterial aggregation, although the exact role of CwpV is yet unknown (Emerson *et al.* 2009b; Reynolds *et al.* 2010).

The presence of antibodies in patient sera to these and other 'minor' surface components suggests firstly, that they are expressed on the bacterial surface during human CDI and secondly, that proteins, other than SlpA, containing CWBD (PFam PF04122) repeats, are likely to play key roles in *C. difficile in vivo* physiology and/or pathogenesis such as cell signaling, matrix degradation and adherence and therefore warrant further study (Drudy *et al.* 2004; Pechine *et al.* 2005a; Wright *et al.* 2008).

1.3 Aims

The aim of the research in this thesis was to understand more about the role of the surface proteins in *C. difficile,* in particular the roles of the S-layer and its 28 paralogs containing the PFam04122 motif.

The work is divided into three main parts:

- Use of a novel gene knockout system for Clostridia, the ClosTron, to obtain mutants and characterise any obtained phenotype(s).
- Establish whether passively administered systemic (circulating) antibodies to selected surface proteins on the bacterium itself can prevent CDI in an animal model.
- Acquire structural data on selected surface layer proteins (SLPs) to gain further insights into their roles in CDI pathogenesis.

The findings presented in this thesis have revealed that certain surface proteins of *C. difficile* play key roles in S-layer biogenesis, adhesion *in vitro* and may alter expression of known virulence factors. The identified surface proteins therefore mark a step forward in CDI research but one that is far from complete. Research into the use of surface proteins to prevent CDI suggests that surface proteins although important, may not be as important as the *C. difficile* toxins or not targetable for CDI intervention via circulating antibodies.

2 General Methods

2.1 General Molecular biology methods

2.1.1 DNA sequencing

Sequencing was performed by Cogenics™ using the ABI 3730xl DNA Analyzer platform (Applied Biosystems).

2.1.2 PCR Primer synthesis

Primers were synthesised by Eurofins MWG Operon (Ebersberg, Germany) by automated Oligo-2000 and ultra parallel HTS synthesisers with phosphoramidite chemistry.

2.1.3 Gene synthesis

Synthetic genes were synthesised by GENEART (Regensburg, Germany) using *E. coli* expression optimisation 'GeneOptimizer® technology (given differences in codon bias between *E. coli* and *C. difficile*) where necessary or by DNA2.0 (California, USA). Synthetic genes were either cloned into a standard vector backbone or chosen/provided vector, for further details see chapter specific methods.

2.1.4 Agarose gel electrophoresis

Agarose gel electrophoresis was performed unless otherwise stated with 1% (wt/vol.) agarose in tris-acetate-EDTA buffer (TAE) plus ethidium bromide (Fluka) or SYBR Safe Stain (Invitrogen), or via primed E-Gel® Pre-cast 0.8% agarose gel(s) on a E-Gel® iBase™ Power System (Invitrogen) as per manufacturer's instructions.

2.1.5 Polymerase Chain Reaction (PCR)

PCR was performed using either illustra PuReTaq Ready-To-Go™ PCR Beads (GE Healthcare) as per manufacturer's instructions, FailSafe Premix E (EPICENTRE Biotechnologies) with Expand High Fidelity^{PLUS} polymerase (Roche) or Platinum Taq (Invitrogen) in a Bio-Rad DNA Engine (PTC-200) Peltier Thermal Cycler. For colony PCR 2.5 µl of selected colony(s) suspension (in 10 µl molecular biology water) was used as template DNA either directly (*E.coli*) or after microwaving for 2 min (*C. difficile* only). Cycling temperatures for primer pairs are mentioned where necessary.

2.1.6 Digestion of DNA

DNA (PCR product(s), plasmid(s) or otherwise) was digested using either standard New England Biolabs (NEB) restriction enzymes or FastDigest® restriction enzymes (Fermentas) as described by the manufacturers. Individual restriction enzyme(s) used in digestion reactions are mentioned where necessary.

2.1.7 Ligation of DNA

DNA was ligated using the Rapid DNA Ligation Kit (Fermentas) or T4 ligase with T4 ligase buffer (NEB), used as described by the respective manufacturers.

2.1.8 Purification of plasmid DNA

Plasmid purification "miniprep" was undertaken using the QIAprep Spin Miniprep Kit (Qiagen) as described by the manufacturer. Eluted DNA was either used directly or frozen at -20 °C.

2.1.9 Purification of PCR product(s)

PCR products were purified using the QIAquick PCR Purification Kit (Qiagen) as described by the manufacturer. Eluted DNA was either used directly or frozen at -20 °C.

2.1.10 Gel extraction of DNA

DNA band(s) were excised from an agarose gel after electrophoresis and purified using the QIAquick gel extraction kit (Qiagen) using protocols provided by the manufacturer. Eluted DNA was either used directly or frozen at -20 °C.

2.1.11 Small scale genomic DNA extraction

Genomic DNA (gDNA) was extracted using the DNeasy Blood and Tissue Kit (Qiagen) via enzymatic lysis using protocols provided by the manufacturer. Eluted gDNA was either used directly or frozen at -20 °C.

2.1.12 RNA extraction

2.1.12.1 Small scale Total RNA extraction

C. difficile was re-streaked on FAA plate(s) and incubated for 24 hr at 37 °C. An appropriate culture volume was inoculated with a streak of colonies and grown overnight at 37 °C. Total RNA was extracted from the *C. difficile* culture using RNeasy mini kit (Qiagen) with enzymatic lysis and proteinase K digestion as per manufacturer's instructions. Total RNA was treated with Turbo DNAse™ (Ambion) as per Section 2.1.12.2.

2.1.12.2 Large scale Total RNA extraction

C. difficile was re-streaked on FAA plate(s) and incubated for 24 hr at 37 °C. An appropriate starter culture volume was inoculated with a streak of colonies and grown overnight at 37 °C. This culture was used to inoculate a final vessel which was then grown until the desired endpoint (24 hr - stationary phase (Chapter 5 Section 5.2.2.1)). At the appropriate endpoint RNAprotect (Qiagen) was added to the culture and incubated at room temperature for 5 min as per manufacturer's instructions. The cells were then pelleted and resuspended in 1.2 ml TRIzol (Invitrogen) and transferred to a Lysing Matrix B (MPBio) ribolyser tube. Samples were ribolysed at 6.5 m/sec for 45 sec then transferred to a tube containing 240 µl chloroform. The tube(s) were vigorously shaken for 15 sec then centrifuged at 12000-15000 g for 10 min. The aqueous phase was removed and added to 600 µl chloroform and the process repeated. The aqueous phase was washed a further time with 600 µl chloroform before transferring to 600 µl isopropanol plus 0.3 M sodium acetate (final concentration) before storing at -80 °C to aid precipitation. The RNA was pelleted at 12000-15000 g for 10 min then washed with 500 µl 70% ethanol then centrifuged as before. The supernatant was removed and the pellet allowed to air dry, before resuspension in 800 µl RNase Free water. To remove DNA contamination RNA was split into two 400 µl aliquots and treated with 20 units TURBO™ DNase (2 x 10 units 30 min apart) for 1 hr. DNase was inactivated by addition of 30 mM EDTA followed by heating to 75 °C for 10 min. Following DNAse treatment 600 µl isopropanol plus 0.3 M sodium acetate (final concentration) were added, mixed and precipitation allowed to proceed at -80 °C for 24 hr. The RNA was

pelleted at 12000-15000 g for 10 min then washed with 500 µl 70% ethanol then centrifuged as before. The supernatant was removed and the pellet allowed to air dry, before resuspension in 200 µl RNase Free water followed by measurement on a Nanodrop 1000. To check for DNA contamination PCR was undertaken with 1 µl RNA sample using gene specific primers. If DNA was present, RNA sample was diluted to 800 µl and DNAse treated as before. The process was repeated until residual DNA contamination was not detectable by PCR. The RNA was purified with RNeasy MinElute Cleanup kit to remove contaminating protein and buffer constituents and eluted with 80 µl RNase free water, followed by measurement by Nanodrop 1000. To remove residual buffer constituents from spin column purification, RNA was precipitated with 3.5 M lithium chloride precipitation solution (final concentration) (Ambion) before re-suspension in RNase free water to ≥200 ng/µl. RNA was then aliquotted and stored at -80 °C.

2.2 ClosTron Gene Knockout system

The ClosTron gene knock-out system developed by Heap *et al.* (Heap *et al.* 2007) was for used for insertional inactivation of the chosen gene(s). Intron integration sites within the chosen genes were identified and intron re-targeting IBS, EBS1d and EBS2. PCR primers were designed using the algorithms available on the TargeTron Design Site (http://www.sigmaaldrich.com) or the implementation of the Peruka (Perutka *et al.* 2004) algorithm available at (http://www.clostron.com/clostron1.php).

The 353 bp region containing the IBS, EBS1d and EBS2 sequences responsible for intron specificity was either: re-targeted by 'splicing-by-overlap extension' (SOEing) PCR using the primers designed above in conjunction with the EBS universal primer (CGAAATTAGAAACTTGCGTTCAGTAAAC) which, after gel purification, was cloned into pMTL007C-E2 via HindIII and BsrGI sites and sequenced *in situ* with primers CspFdx-F1 (GATGTAGATAGGATAATAGAATCCATAGAAAATATAGG) and pMTL007-R1 (AGGGTATCCCCAGTTAGTGTTAAGTCTTGG). Alternatively, the 353 bp region was synthetically synthesised based on *in silico* mutation of an intron (re)targeting

region template (Figure 2.1) using primers designed by the relevant intron retargeting site.

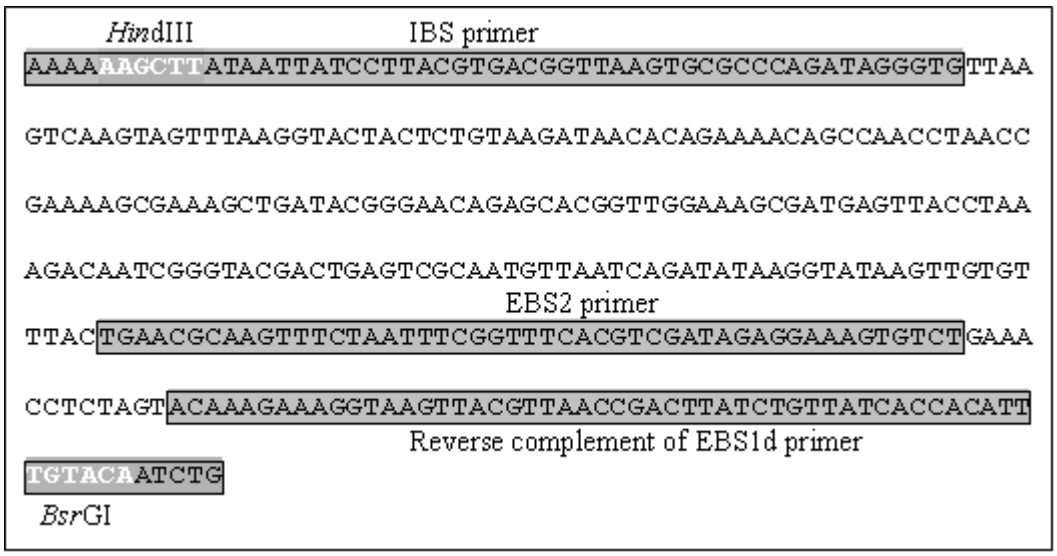


Figure 2.1 The intron re-targeting region from pMTL007. The unique HindIII and *Bsr*GI restriction sites are shown in white.

The derivative pMTL007 plasmid construct(s) were transformed, via electroporation, into the conjugative donor *E. coli* CA434 and then transferred, via conjugation, into *C. difficile* 630ΔErm (Purdy *et al.* 2002).

Successful transconjugants and subsequent integrants were selected for by streaking onto *C. difficile* selective media (E&O Labs) with the addition of thiamphenicol (15 µg/ml) and then on sBHI agar containing erythromycin (5 µg/ml), respectively. Confirmation of knockout was performed using PCR with primers flanking the intron insertion site (gene specific primers mentioned in chapter specific methods) and using the Erm-RAM primers (ErmRAM_F ACGCGTTATATTGATAAAAATAATAATAGTGGG and ErmRAM_R ACGCGTGCGACTCATAGAATTATTTCCTCCCG). Erythromycin mutants were both re-streaked onto sBHI agar containing erythromycin (5 µg/ml) and spread onto FAA plus 5% horse blood (Oxoid) for safekeeping.

2.3 Protein expression

2.3.1 'Harvesting' of bacterial cultures

Bacterial cultures were centrifuged at 3488 g for 20 min (culture volume <250 ml) or 4652 g for 40 min (culture volume >250 ml) unless otherwise stated.

2.3.2 Small scale expression trials

An appropriate *E. coli* host was transformed with an expression vector containing a gene of choice and grown overnight at 37 °C on L-agar plus appropriate antibiotic. One colony was inoculated into 2 ml L-broth plus appropriate antibiotic plus 0.5% glucose and grown until OD_{600} ~0.6. Cultures were then cooled to 30 or 16 °C then induced with 1 mM IPTG and grown for a further 4 hr or 16 hr respectively, before harvest. Soluble/insoluble expression analysis was performed by resuspending the pellet in Cellytic B reagent™ (Sigma) plus Benzonaze™ and incubating for 5 min at room temperature. Cell suspension was then centrifuged at ≥10,000 g for 5 min to pellet insoluble material. The soluble fraction (supernatant) was mixed with 4x LDS sample buffer, while insoluble material was resuspended in 1x LDS loading buffer, ready for SDS-PAGE.

2.3.3 Large scale expression

Large scale expression was similar to small scale expression (Section 2.3.2) except after plating, a 50 ml starter culture of terrific broth (TB) with appropriate antibiotic plus 0.5% glucose was inoculated and grown overnight at 30 °C with shaking. The starter culture was then inoculated into 950 ml of the aforementioned supplemented TB media and grown until OD_{600} ~0.6. Cultures were then cooled to 30 or 16 °C then induced with 1 mM IPTG and grown for a further 4 hr or 16 hr respectively before harvest. Cell pellets were either used directly or frozen at -20 °C.

2.3.3.1 Protein extraction/Cell lysis

Bacterial cells from large scale expression were thawed on ice and resuspended in chromatography binding/wash buffer. Cells were then sonicated for 6 x 30 sec (with 30 sec cooling) before centrifugation at 47,860 g (20,000 rpm Sorval SS-24 rotor) to

remove cell debris. The soluble fraction was decanted and retained on ice until processing, insoluble fraction was used for refolding if necessary.

2.4 Protein Purification

2.4.1 Affinity Chromatography

2.4.1.1 Glutathione sepharose affinity purification

GST tagged proteins were purified using a GSTrap™ column (GE Healthcare)

Binding/Wash buffer – 1 x PBS (±10 mM DTT) pH 7.3

Elution buffer - 50 mM tris-HCl, 10-50 mM reduced glutathione (±10 mM DTT) pH 8.0

Alterations to the above buffers, along with flow rates, gradients and fraction sizes are mentioned where necessary.

2.4.1.2 Purification of IgG by affinity purification

IgG fraction of animal serum was purified using either HiTrap™ Protein A HP (guinea pig serum) or Protein G (sheep serum)

Binding/Wash buffer – 20 mM sodium phosphate pH 7.0

Elution buffers - 0.2 M sodium citrate pH 3.0 (Protein A) or 0.1 M glycine pH 2.7 (Protein G)

Neutralisation buffer - 1 M tris pH 9.0

2.4.1.3 Immobilised metal-ion affinity chromatography (IMAC)

Native IMAC was performed on an ÄKTA Design FPLC (GE Healthcare) using either:

HiTrap™ chelating HP column (GE healthcare) charged with 0.1 M nickel sulphate or 0.1 M cobalt sulphate (Sigma) or pre-charged HisTrap™ column

Binding/Wash buffer- 50 mM tris, 0.5 M NaCl, 20 mM imidazole pH 8.0

Elution buffer- 50 mM tris, 0.5 M NaCl, 0.5 M imidazole pH 8.0

Or using BD TALON™ (Clontech) resin (recharged where necessary with 0.1 M cobalt chloride)

Binding/Wash buffer- 50 mM sodium phosphate, 300 mM NaCl pH 7.0

Elution buffer- 50 mM sodium phosphate, 300 mM NaCl, 150 mM imidazole pH 7.0.

Alterations to the above buffers, along with flow rates, gradients and fraction sizes are mentioned where necessary.

2.4.2 Ion-exchange chromatography (IEX)

Anion Exchange Chromatography was performed primarily using Q-sepharose™ fastflow (FF) (GE healthcare). Where MonoQ beads were used, for higher resolution, buffers remained the same.

Binding/Wash buffer – 50 mM tris pH 8.0

Elution buffer 50 mM tris, 1 M NaCl pH 8.0

Cation Exchange Chromatography was performed primarily using S-sepharose™ FF

Binding/Wash buffer – 50 mM sodium acetate pH 6.0

Elution buffer - 50 mM sodium acetate, 1 M NaCl, pH 6.0

Alterations to the above buffers, along with flow rates, gradients and fraction sizes are mentioned where necessary.

2.4.3 Hydrophobic Interaction Chromatography (HIC)

HIC scouting experiments were performed using 1 ml HiTrap™ HIC selection kit (GE healthcare). Large scale HIC was undertaken using 5 ml HiTrap phenyl HP (GE Healthcare) columns.

Trial buffers were:

Binding/Wash buffer – 50 mM sodium phosphate, 1 M ammonium sulphate pH 7.0

Elution buffer- 50 mM sodium phosphate pH 7.0

Alterations to the above buffers, along with flow rates, gradients and fraction sizes are mentioned where necessary.

2.5 Protein analysis

2.5.1 Sodium dodecyl sulphate- Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Denaturing SDS-PAGE was carried out using NuPAGE® Novex® bis-tris 12well or 10well 1.0 mm pre-cast gels (Invitrogen), NuPAGE LDS 4X LDS sample buffer (Invitrogen) and either stained with SimplyBlue™ Safe Stain (Invitrogen) or with Pierce Silver Stain Kit II (Pierce, Thermo) as per manufacturer's instructions.

2.5.2 Blue Native Polyacrylamide Gel Electrophoresis (BN PAGE)

Blue Native PAGE was performed using the NativePAGE™ bis-tris gel system (Invitrogen) as per manufacturer's instructions. Samples were prepared with NativePAGE™ sample buffer (4X) and ran on either NativePAGE™ Novex® 3-12% or 4-16% bis-tris 1.0 mm pre-cast gel(s), using 200 ml light blue cathode buffer (containing 0.002% Coomassie G-250) (1 ml NativePAGE™ cathode additive (20X) plus 10 ml NativePAGE™ running buffer (20X)) together with 1X NativePAGE™ anode buffer.

2.5.3 Western blotting

SDS-PAGE gel(s) were transferred onto nitrocellulose or PVDF membrane(s) using the XCell SureLock® Mini-Cell and XCell II™ Blot Module Kit with NuPAGE® Transfer Buffer. Membranes were blocked with 5% w/v non-fat dried milk (Fluka) in water or PBS plus 0.1% Tween 20 (PBS/T) (blocking buffer). Incubation with primary antibody (at appropriate dilution in blocking buffer) was for 1 hr followed by 3 washes in PBS/T. If necessary, a further incubation with secondary antibody again for 1 hr was followed by 3 washes in PBS/T prior to chemiluminescent detection/development using ECL™ western blotting detection reagents (Amersham and/or Pierce).

Common antibodies and dilutions used in western blot:

Mouse monoclonal T7•Tag® HRP Conjugate (Novagen) - 1:5000

Mouse Monoclonal Anti-PolyHistidine (His_6) −Peroxidase (Sigma) – 1:7000

2.5.4 Mass spectrometric identification of proteins

These analyses were conducted at the Centre for Proteomic Research, University of Southampton. Excised protein bands were subjected to in situ trypsin digestion using previously described methods (Shevchenko *et al.* 1996). The resulting

peptides were separated by nano-reverse phase liquid chromatography, using a Dionex C18 PepMap, 3 µm, 100 Å (150 mm x 75 µm, i.d.) column, and electrosprayed into a quadrupole time-of-flight tandem mass spectrometer. All data were acquired using a Q-tof Global Ultima (Waters Ltd) fitted with a nanoLockSpray™ source. A survey scan was acquired from m/z 375 to 1800 with the switching criteria for MS to MS/MS including ion intensity and charge state. The collision energy used to perform MS/MS was varied according to the mass and charge state of the eluting peptide. All MS/MS spectra were automatically processed using MASCOT distiller (Matrix Science) and subsequently searched against a FASTA formatted listing of protein sequences predicted from the *C. difficile* 630 genome sequences using the search algorithm MASCOT ver2.2. The following parameters were used: parent mass tolerance was 150 ppm, fragment mass tolerance 0.25 Da, carbamidomethylation was set as a fixed modification, the oxidation of methionine as a variable modification and a maximum of one missed cleavage was allowed. The significance threshold for search results was set at $p < 0.05$ which indicates identity or extensive homology.

2.6 Clostridium difficile strains and growth conditions

The *C. difficile* strains used throughout this thesis are summarised in Table 2.1. Liquid cultures were grown in pre-reduced supplemented brain-heart infusion broth (sBHI) (per litre of broth – 37 g brain heart infusion (Oxoid), 5 g yeast extract (Oxoid) 1 ml resazurin (1.0 mg/ml) 0.5 g L-cysteine HCl, 10 ml haemin solution, 1 ml vitamin K1 solution) at 37 °C in an atmosphere of 10% H_2, 10% CO_2 and 80% N_2 in a Don Whitley MG1000 anaerobic workstation.

Plate cultures, unless otherwise stated, were grown on pre-reduced Fastidious Anaerobe Agar (FAA) plus 5% horse blood (Oxoid) and incubated anaerobically as above.

2.7 Other general methods

2.7.1 Toxin A or B ELISA

Affinity purified sheep anti-toxoid A and B IgG was biotinylated using EZ-Link® Sulfo-NHS-Biotin reagents (Pierce) as per manufacturer's instructions. Prior to and after

biotinylation IgG was dialysed against 50 mM HEPES, 0.15 M NaCl pH 7.4 overnight at 4 °C. Unbiotinylated sheep anti-toxoid A and B IgG was coated onto a Maxisorb plate at 5 µg/ml in PBS and incubated for 1 hr or overnight at 4 °C. Plates were washed three times with PBS plus 0.1% Tween 20 (Sigma) (PBS/T) followed by application of PBS/T plus 5% foetal calf serum (blocking buffer) and incubated for 1 hr. Plates were washed a further three times before the application of appropriate samples and toxin A and B standards (diluted to 1000 ng/ml) and incubated for 1 hr. Plates were washed a further three times before application of either biotinylated sheep anti-toxoid A (1 µg/ml) or B IgG (10 µg/ml) and incubated for 1 hr. Plates were washed as before then streptavidin-HRP conjugate (GE Healthcare) (1:1000, in blocking buffer) applied and incubated for 10 min followed by a final three washes. Detection was undertaken with the addition of a TMB substrate and TMB stop solution (BioFX), then read at 450 nm using a Tecan™ Sunrise microtitre plate reader. Toxin concentrations were estimated using a sigmoidal four-parameter logistic standard curve.

C. difficile culture supernatants for toxin A or B ELISA were obtained by removing bacterial cells from *C. difficile* culture by centrifugation, then dialysing the culture supernatant for 24 hr into 50 mM HEPES 0.15 M NaCl pH 7.4, using 12 kDa MWCO dialysis tubing before storage at -80 °C if not used directly.

2.7.2 Surface protein extraction

Surface protein extraction was essentially as described by (Cerquetti *et al.* 2000; Calabi *et al.* 2001) using low pH glycine extraction. Briefly, the harvested cell pellets were washed once with 10 mM HEPES containing 100 mM NaCl pH 7.4, re-suspended in 0.2 M glycine pH 2.2 and incubated for 30 min at 22 °C. Cells were then removed by centrifugation and the supernatant neutralised with 2 M tris-HCl pH 9.0. Extracted SLPs were dialysed into 50 mM HEPES containing 0.15 M NaCl pH 7.4 and stored at -80 °C if not used directly.

Table 2.1 *C. difficile* strains used in this thesis

Referred to in this text	Ribotype	Toxinotype	Toxin production	Source	ARL Cardiff Reference	ATCC Reference
R20291	027	III	A+B+	Stoke Mandeville Hospital	R20291	13366
VPI10463	001	0	A+B+	Human abdominal wound isolate		43255
ARL001	001	0	A+B+	ARL	R8652	11209
630ΔErm	630	0	A+B+	A gift from N. Minton (Nottingham University) as generated in Hussain *et al.* (2005)		

CHAPTER 3
Cwp84

3 CHAPTER 3 – Cwp84

3.1 Introduction

3.1.1.1 Proteases in infection

Clostridia exist primarily as apparently harmless saprophytes and are normally found in anaerobic environments high in fermentable organic compounds including soils, aquatic sediments and the intestinal tracts of animals. Being saprophytic, Clostridia produce a wide variety of cellular and extracellular enzymes to degrade nutrients and large biological molecules e.g. proteins, lipids, collagen, cellulose. This action normally contributes to the natural cycles in the environment, however upon an infection (usually opportunistic) these factors play a role in invasion and pathology (Allen *et al.* 2003; Todar 2008).

Directly or indirectly proteases therefore play key roles during the colonisation process as adhesins, as part of nutrient acquisition or in the processing of bacterial proteins necessary for growth and ensuring continued proliferation (Maeda 1996; Potempa & Pike 2005). For example, the pathogenesis of *C. difficile* is primarily due its toxins (TcdA and TcdB), both of which are automatured via cysteine protease activity, a mechanism shared by *Vibrio cholerae* toxins (Sheahan *et al.* 2007; Giesemann *et al.* 2008) highlighting how, through the action of maturing these toxins, cysteine proteases promote infection indirectly.

3.1.1.2 Cysteine proteases

Cysteine proteases are so called as the sulphydryl group of a cysteine residue acts as a nucleophile allowing hydrolysis of the peptide bond. The nucleophilic cysteine thiol acts with a histidine forming a catalytic dyad. Orientation of the imidazole ring of the active histidine and formation of an 'oxyanion hole' is believed to be assisted by a glutamine preceding the cysteine and an asparagine following the catalytic histidine respectively. The inclusion of either of these additional residues therefore leads to the active site residues also being described as a catalytic triad, although only the Cysteine and Histidine participate catalysis (Otto & Schirmeister 1997; Rawlings *et al.* 2010). The structure of papain (PDB: 9PAP) demonstrates a globular

protein with two interacting domains, with substrates bound in the cleft of the domains into the 'oxyanion hole' (Drenth *et al.* 1968) (Figure 3.1.1).

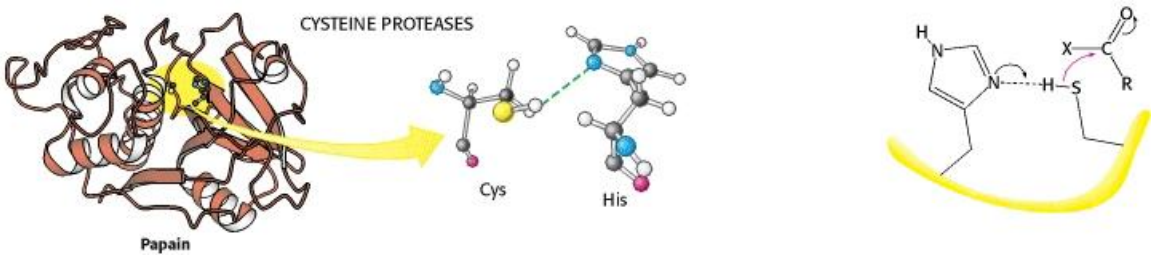


Figure 3.1.1 Cysteine proteases and their active sites. (A) Cysteine protease papain and highlight of nucleophilic cysteine thiol and active histidine. (B) The cleavage of the peptide carbonyl group by the cysteine and histidine. Taken from Berg *et al.* (2002)

Like many proteins the tertiary structure is more conserved than the primary sequence, hence cysteine proteases can be divided into clans and further sub-divided into families on the basis of the architecture of their catalytic dyad or triad and on their origin e.g. C1 papain, C3 and 4 viral (Rawlings & Barrett 1993; Barrett & Rawlings 2001).

Cysteine proteases, like other proteases, are involved in pathogenesis in a variety of ways, for example papain-like, type III secreted, cysteine proteases of *Yersinia* that arrest the cell cycle (Yao *et al.* 2009), the degradation of complement by *Prevotella intermedia* (Potempa *et al.* 2009) or surface-associated proteolytic degradation of IgA in *Entamoeba histolytica* (Garcia-Nieto *et al.* 2008).

3.1.1.3 Discovery and characterisation of a surface located cysteine protease in *C. difficile*.

A putative surface located thiol protease was first discovered by Seddon & Borriello (1992). The cell surface associated protease exhibited trypsin-like substrate specificity but was inhibited by thiol protease inhibitors, with an optimum pH of 7.5 and temperature of 37 °C. Seddon & Borriello (1992) therefore suggested that the *C. difficile* thiol protease was more like clostripain (C11 family cysteine protease),

which cleaves the carboxy peptide bond of arginine, than trypsin (which cleaves the carboxyl side of the lysine or arginine). They also found there was no correlation between virulence and protease production, suggesting that the protease was not related to virulence.

Following the examination of a 37 kbp DNA fragment flanking *slpA* (gene of main S-layer) in the genome of *C. difficile* 630, Karjalainen *et al.* (2001) and Calabi *et al.* (2001) found 17 open reading frames (ORFs) in the same genetic locus as *slpA*, with a further 11 suggest to exist elsewhere in the *C. difficile* genome. Eleven of which encoded a domain present at either the N- or C-terminus with homology to the N-terminal cell wall-anchoring domain of the CwlB autolysin of *B. subtilis* (Kuroda & Sekiguchi 1991). Subsequently, 28 genes encoding an N- or C-terminal cell wall binding domain (CWBD (PFam04122)) have been found (Sebaihia *et al.* 2006). However, only two genes contain a conserved peptidase C1 (papain) family domain at the N-terminus (and three CWBDs at the C-terminus) (Figure 3.1.2), *cwp84* (*cd2787*) found in the 37 kbp putative surface protein cluster and named from the putatively 84 kDa protein the gene encodes (Savariau-Lacomme *et al.* 2003), and *cd1751*, less attractively named from the *C. difficile* 630 genome gene numbering system, found a large distance away from *slpA* the *C. difficile* genome.

803 aa (87,280Da predicted)

Signal peptide 1-32

Peptidase_C1A.....104-322

3x CWBD 507-595, 604-697, 702-793

Active residues (Q110, C116, H262, N294)

Figure 3.1.2 Domain structure of Cwp84. Sig – signal peptide, Peptidase_C1 - C1A (papain) peptidase family, CWBD – cell wall binding domain (Pfam04122). Flags denote Cwp84 active residues C116 and H262. Putative domain location(s) obtained from *C. difficile* 630 genome annotations, available at ncbi.nlm.nih.gov. [Note: domain structure of CD1751 is similar to Cwp84]

To understand more about *cwp84*, expression and analysis of the *cwp84* gene was undertaken. Savariau-Lacomme *et al.* (2003) found the *cwp84* gene is highly conserved and appears to be transcribed as a monocistronic message particularly prominent at early exponential growth, while amino acid sequence analysis suggested the conserved active site residues of C116, H262 and N287. Initial proteolytic assays using recombinant Cwp84 did not confirm specific substrates, nor did typical cysteine protease inhibitors inhibit the recombinant Cwp84 produced. Interestingly, free GST-Cwp84 N-term fragments found during expression of the whole *cwp84* ORF (including signal peptide) in *E. coli* lead to the suggestion that Cwp84 may auto-cleave (auto-mature), a mechanism found in other cysteine proteases e.g. papain (Vernet *et al.* 1995). Savariau-Lacomme *et al.* (2003) also found no correlation between expression of *cwp84* and virulence which, taken with the above data, led to the suggestion that Cwp84 was key in the physiology of *C. difficile* possibly in cleavage of other cell surface-associated proteins e.g. SlpA.

By expressing only the cysteine protease domain (343 aa total) (less signal peptide) tagged to GST, Janoir *et al.* (2004) found that the proteolytic activity on gelatine and BAPNA (N-α-Benzoyl-DL-arginine 4-nitroanilide hydrochloride, a chromogenic trypsin substrate) was inhibited by cysteine and serine protease inhibitors. The refolded recombinant Cwp84 displayed a molecular weight of 45 kDa on SDS-PAGE and proved to have a weaker activity than trypsin but higher than papain. The proteolytic activity appeared to decrease quickly, suggesting enzyme instability. A significant finding by Janoir *et al.* (2004) was that CDI patient serum contained antibodies to the cysteine protease domain of Cwp84. This highlights how an immune response is generated to Cwp84 but also how Cwp84 must be accessible to the immune system to generate the immune response i.e. it is surface located. Cwp84's surface-based location and immunogenicity was confirmed by both Pechine *et al.* (2005a) who found only 2 out of 17 patients tested did not have antibodies to Cwp84 and Wright *et al.* (2008) who found antibodies to Cwp84 in patient serum by proteomic analysis of reacting surface proteins.

Full length recombinant His_6 tagged Cwp84 (rCwp84) displayed an inhibition profile similar to that seen previously, with typically 10% activity remaining after serine and cysteine protease inhibitor treatment, except when 100 µM E64 was used which fully inhibits Cwp84. rCwp84 appeared to be automatured upon exposure to DTT, presenting as a final 'mature' 61 kDa species (mCwp84). However, both 'immature' (rCwp84) and 'mature' (mCwp84) displayed proteolytic activity. In support of a host tissue and infection dissemination hypothesis, Cwp84 was found to degrade fibronectin, laminin, and vitronectin but not type IV collagen. Moreover, as found in Seddon & Borriello (1992) the pH optimum was pH 7.5 (Janoir *et al.* 2007).

3.1.1.4 Localisation of Cwp84

Use of mCwp84 to generate antiserum led to the identification of Cwp84 in various cell fractions of *C. difficile*. Cwp84 was found as a 90 kDa pre-proprotein (including signal peptide) in the cytoplasm, an 82 kDa protein in the membrane and cell wall fractions and finally as a putatively glycosylated 85 kDa version in low pH glycine surface protein extracts. A small amount of 61 kDa protein was also found in the glycine extract (Janoir *et al.* 2007).

Proteomic analysis of cell wall and low pH surface extracts only identified Cwp84 as a small fragment in the lysozyme cell wall extract indicating that either Cwp84 is an easily degraded or labile protein, or that it is not found in glycine extracts under the conditions tested (Wright *et al.* 2005).

3.1.1.5 Understanding the role of Cwp84 in *C. difficile*

From gene expression studies it appears that certain surface located genes e.g. *cd2791*, *slpA*, *cwp66,* and especially *cwp84,* are upregulated during environmental stress e.g. high osmolarity, iron depletion or when *C. difficile* is exposed to sub-MIC levels of antibiotics (Deneve *et al.* 2008; Emerson *et al.* 2008; Deneve *et al.* 2009a). In response to fluoroquinolones, the amounts of Cwp84 found on the cell surface did not correlate perfectly with the increased expression of *cwp84*, suggesting that Cwp84 may also be secreted (potentially similar to the supernatant based

proteolytic activity originally eluded to by Seddon & Borriello (1992)) (Deneve *et al.* 2009a).

The hypothesised role(s) of Cwp84 *in vivo* are therefore; an infection dissemination hypothesis, based upon the ability of recombinant Cwp84 to degrade ECM proteins and potential being secreted or Cwp84 playing a key role in the bacterium's physiological response particularly to stress e.g. environmental or chemical (antibiotics) from expression data.

Presented here and reported in Kirby *et al.* (2009), the ClosTron gene knockout system was used to further understand the role of Cwp84. Knockout of Cwp84 results in a bacterial phenotype in which only immature, single chain SlpA is presented at the cell surface. Furthermore, Cwp84 appears not to be an essential virulence factor and bacteria expressing immature SlpA are able to cause disease. Recently, it has also been shown that by inhibition of surface protease activity and use of activity based probes, that it is Cwp84 that mediates SlpA maturation *in vivo*. Dual expression also demonstrated Cwp84 cleavage of SlpA at a defined cleavage site (Dang *et al.* 2010).

Taken together these data provide convincing evidence that Cwp84 plays a key role in S-layer maturation. Corroborating earlier reports, Cwp84 and the lack thereof is not directly correlated with virulence; however a connection between its ability to be up or down-regulated and virulence is not precluded.

3.2 Chapter 3 Specific Methods

3.2.1 Cloning and expression of recombinant Cwp84

3.2.1.1 Genetic methods

3.2.1.1.1 Synthetic synthesis of C. difficile 027 (R20291) strain cwp84 gene

The *C. difficile* 630 *cwp84* gene sequence was BLAST searched against the Sanger Institute's *C. difficile* strain R20291 (Stoke Mandeville, UK) assembly (March, 2008). The resulting match was reverse transcribed and translated *in silico*. The resulting amino acid sequence, flanked by BamHI, two C-term Stop codons and XhoI restriction sites was submitted to GENEART (Regensburg, Germany) for optimised gene synthesis and direct insertion into pET28(a) using the same two restriction sites. The resulting *cwp84* gene is herein referred to as *cwp84*(g).

3.2.1.1.2 Primers used to clone and mutate cwp84

The primers given in Table 3.2.1 were used to clone or mutate *cwp84* from gDNA of *C. difficile* 630ΔErm or *cwp84*(g) directly in pET28a. Resulting PCR products were purified as described in General Methods 2.1.9, then digested (where necessary) with appropriate restriction enzymes (Fast Digest, Fermentas) (Section 2.1.6). The resulting cloned *cwp84* gene(s) are herein referred to as *cwp84*([derived strain]) e.g. *cwp84*(630)

Table 3.2.1 Primers used to clone or mutate rCwp84 and rCwp84$_{33-497}$.

Function	Name	Sequence	Comments
Amplify Native full length *cwp84* (minus signal sequence	cwp84_F	GTG AAT TCC TAT TTT CCT AAA AGA GTA T	As used in Janoir et al. (2007)
	cwp84_R	TGA GCT AGC GCA GAA AAC CAT AAA ACT CTA GAT G	
Change cwp84 (g) reading frame to fit pET b series vectors	cwp84_b_for	GCG GAT CCA CAC AAA ACC	Alter optmised Cwp84 retaining 5' BamHI & 3' XhoI
	cwp84(b)_short_r	GTG CTC GAG TTA TTA TTT GC	
Remove stop codon from cwp84 (g)	cwp84(b)_shrt_fo	GCG GAT CCC ACA AAA CCC	Alter optmised Cwp84 retaining 5' BamHI & 3' XhoI
	cwp84RemStop_rev	CTC CTC GAG TTT GCC CAG CAG	
Change *cwp84* (g) to b reading frame & remove stop	cwp84_b_for	GCG GAT CCA CAC AAA ACC	Alter optmised Cwp84 retaining 5' BamHI & 3' XhoI
	cwp84RemStop_rev	CTC CTC GAG TTT GCC CAG CAG	
Amplification of C. difficile 630 *cwp84* cysteine protease domain	NheIcwp84630_F	GAG GCT AGC GCA GAA AAC CAT AAA ACT CT	5' NheI & 3' EcoRI
	EcoRIcwp84630_R	GAG GAA TTC ATT TGC TGT GTT GTC TAT AAC	
Amplification of cwp84 (g) for TA cloning into pET SUMO vector	Cwp84SUMO_F	AGC CAC AAA ACC CTG GAT GG	Addition of N-term S, C-term stop
	Cwp84SUMO_R	TTA GCT GGT TTT GGT GAT CGC	
Amplification of cysteine protease domain of *cwp84* (g) to fit pTAC-MAT-2	84(AS)HindIII_F	GAG CTC AAG CTT CAC AAA ACC CTG	5' HindIII & 3' EcoRI
	84(AS)EcoRI_R	GAG CTC GAA TTC GCT GGT TTT GG	

3.2.1.1.3 Insertion of cwp84 *into a range of expression vectors*

In order to assess the expression and solubility of recombinant Cwp84, a number of expression vectors were assessed (Table 3.2.2).

Table 3.2.2 Expression vectors used in this chapter

Vector Name	Supplier	Antibiotic Resistance	Tag	5' Restriction site	3' Restriction site
pET28a	Novagen	kan	N-term His_6 N-term His_6.T7	NheI BamHI	EcoRI XhoI
pET24b	Novagen	kan	N-term T7 C-term His_6	BamHI	XhoI
pET43.1a	Novagen	amp	N-term NusA.His_6.S tag	BamHI	XhoI
pET41a	Novagen	kan	N-term GST.His_6.S tag	BamHI	XhoI
pET41b	Novagen	kan	N-term GST.His_6.S tag C-term His_6	BamHI	XhoI
pGEX-6P-1	GE Lifescience	amp	N-term GST	BamHI	XhoI
pTAC-MAT2	Sigma	amp	C-term His_6	HindIII	EcoRI

3.2.1.2 Expression of recombinant Cwp84 variants

3.2.1.2.1 Expression hosts

In order to assess the expression and solubility of recombinant Cwp84, a number of *E. coli* expression hosts were tried (Table 3.2.3)

Table 3.2.3 *E. coli* expression hosts used in this chapter

Name	Supplier	Notes
BL21	Novagen	General purpose expression host
BL21 (DE3)	Novagen	General purpose expression host. Expression of SUMO fusion.
BL21 (DE3) Star	Invitrogen	General purpose expression host with reduced mRNA degradation. For expression of T7 driven promoters
Origami 2 (DE3)	Novagen	Thioredoxin reductase (trxB) and glutathione reductase (gor) mutants for increased cytplasmic disulphide bond formation
Overexpress C41 &C43 (DE3)	Lucigen	BL21(DE3) derivatives empirically selected for toxic gene expression
T7 Express lysY/Iq (High Efficiency)	NEB	Tighter control of expression and less susceptibility to lysis during induction

3.2.1.2.2 General expression protocol for full length pET constructs

Constructs were transformed into the expression strain via heat shock and spread onto L-agar containing the appropriate antibiotic then incubated overnight at 37 °C. Due to plasmid or DE3 lysogen loss, the plating method of induction was used. Two bacterial lawns were required per litre of culture thus two colonies were resuspended in 2 x 100 µl PBS then re-spread onto L-agar and incubated as before. 1 litre of L-broth plus the appropriate antibiotic supplemented with 0.5% w/v D-glucose was inoculated by scraping off and re-suspending the 2 plates of bacterial lawn. Cultures were grown at 30 °C until OD_{600} ~0.6. Cultures were then cooled to 30 °C then induced with 1 mM IPTG for 1hour or cooled to 16 °C then induced overnight. Cells were then harvested and frozen at -20 °C

3.2.1.2.3 General expression protocol for other constructs

All other construct expression was performed as detailed in General Methods (small scale- section 2.3.2 or large scale- section 2.3.3)

3.2.1.3 Purification of recombinant Cwp84

3.2.1.3.1 Purification of soluble His_6 or MAT tagged rCwp84

His_6 or MAT tagged material was purified in native IMAC as described in General Methods (Section 2.4.1.3).

3.2.1.3.2 Purification and refolding of insoluble His tagged rCwp84

Isolation of insoluble material was based on the inclusion body isolation procotol found in the Pierce Protein Refolding Kit (Thermo). Frozen cell pellets were thawed and resuspended in lysis/wash solution containing 50 mM tris, 2% triton X-100, 5 mM DTT, 500 mM NaCl pH 8.0 and sonicated for 6x30 sec on ice. Cell debris and insoluble material was recovered by centrifugation at 38,770 g (18,000 rpm Sorval SS-34 rotor) for 20 min. The pellet was directly resuspended in ice cold 50 mM tris, 7 M guanidine hydrochloride (GnHCl) pH 8.0 and left overnight at 4 °C with stirring to solubise. After solubisation the suspension was centrifuged for 30 min at 47,860 g at 4 °C, the supernatant was then filtered with a 0.2 µm filter and returned to ice ready for denaturing IMAC.

Denaturing IMAC was undertaken on a 1 ml HisTrap™ column with 50 mM tris, 6 M GnHCl, 20 mM imidazole, pH 8.0 (binding/wash buffer) and as above but with 0.5 M Immidazole in the elution buffer. Fractions were analysed by SDS-PAGE by first removing GnHCl using Pierce SDS-PAGE Sample Prep Kit (Pierce) as per manufacturer's instructions, followed by silver staining and western blot where necessary.

Fractions of denaturing IMAC purified His.T7.Cwp84 were refolded as outlined in the Pierce Protein Refolding Kit (Thermo), using the low protein concentration protocol. Refolding was allowed to proceed overnight at 4 °C. Assessment of refolding on protein was performed by SDS-PAGE.

3.2.1.3.3 Purification of (pGEX) GST tagged rCwp84

GST tagged protein was purified using a glutathione column as described in General Methods (Section 2.4.1.1).

3.2.1.3.4 Purification of SUMO tagged rCwp84$_{33-497}$

Expression was performed as described in general expression protocol (Section 2.3.3) using *E. coli* BL21 (DE3). His_6.SUMO tagged proteins were first purified by IMAC (Section 2.4.1.3) then dialysed into SUMO protease cleavage buffer (50 mM tris-HCl, 150 mM NaCl pH 8.0, 0.2% igepal, 1 mM DTT). A second round of IMAC was performed using 50 mM tris-HCl, 150 mM NaCl pH 8.0 as binding/wash buffer, to ensure any uncleaved material and/or contaminants bound, with cleaved untagged protein not binding. To remove residual detergent from cleavage buffer, the flow through was dialysed into IEX (Q pH 8.0) start buffer and purification undertaken as per General Methods (Section 2.4.2).

3.2.1.4 *In vitro* transcription and translation (IVT) of recombinant Cwp84 and small scale batch purification of Cwp84

In vitro transcription and translation (IVT) was undertaken using EasyXpress Protein Synthesis Kit (Qiagen) as per manufacturer's instructions, using plasmid miniprep as template DNA. Expression analysis was performed using 1 µl of the IVT reaction in SDS-PAGE, while the remainder (49 µl) was used in native, small scale, batch purification with BD TALON (Clontech) IMAC resin as per the manufacturers protocol "Appendix B". Briefly, 100 µl resin slurry was first equilibrated with 2 washes in binding/wash buffer followed by re-suspension and incubation with IVT reaction for 10 min at room temperature. Unbound material was removed with a further 2 washes with binding/wash buffer. Bound material was then eluted, twice, by resuspending resin in elution buffer. Fractions were analysed by SDS-PAGE and western blot with appropriate antibodies.

3.2.1.5 Treatment of purified rCwp84 with DTT

Insoluble refolded rCwp84 was incubated overnight with 5 mM DTT at 4 °C. Soluble material eluted from IMAC was either incubated with 5 mM DTT at room temperature or at 37 °C for 4 hr.

3.2.1.6 Treatment of purified rCwp84 with thrombin

Purified protein from pET28a constructs was incubated with thrombin as per small scale optimisation protocol with Thrombin, Restriction grade (Novagen).

3.2.2 Knockout of cwp84

3.2.2.1 Genetic methods

3.2.2.1.1 Knockout of Cwp84 – ClosTron system

The ClosTron gene knock-out system was used for insertional inactivation of *cwp84* using protocols outlined in General Methods (Section 2.2). Retargeting primers were designed by the TargeTron site (Sigma) (Table 3.2.4). Retargeting for insertion site 347|348a was performed by SOE PCR, while other sites retargeting 353 bp region was synthetically synthesised.

Flanking primers used to confirm successful integration into *cwp84* were (Cwp84_F TGAGCTAGCGCAGAAAACCATAAAACTCTAGATG & Cwp84_R ATACTCTTTTAGGAAAATAGGAATTCAC)(Janoir *et al.* 2007).

3.2.2.1.2 Genomic DNA extraction

Small scale genomic DNA extraction was performed as per General Method's (Section 2.1.11).

3.2.2.1.3 Analysis of the CD1751 gene

Full length *cd1751* was amplified by PCR using primers in Chapter 5 Table 5.2.2.

Table 3.2.4 IBS, IBD1d and EBS primers for retargeting ClosTron Ll.ltrB intron

Gene	insertion location	IBS	IBS1d	EBS2
cwp84	68|69a	AAAAAAGCTTATAA TTATCCTTAGGTATC GTTGACGTGCGCCC AGATAGGGTG	CAGATTGTACAAATG TGGTGATAACAGAT AAGTCGTTGACACTA ACTTACCTTTCTTTGT	TGAACGCAAGTTTCT AATTTCGATTATACCT CGATAGAGGAAAGT GTCT
cwp84	347|348a	AAAAAAGCTTATAA TTATCCTTAGAAAA CGACCAGGTGCGCC CAGATAGGGTG	CAGATTGTACAAATG TGGTGATAACAGAT AAGTCGACCAGCAT AACTTACCTTTCTTTG T	TGAACGCAAGTTTCT AATTTCGATTTTTTCT CGATAGAGGAAAGT GTCT
cwp84	677|678a	AAAAAAGCTTATAA TTATCCTTATATTGC ATTATAGTGCGCCC AGATAGGGTG	CAGATTGTACAAATG TGGTGATAACAGAT AAGTCATTATAGCTA ACTTACCTTTCTTTGT	TGAACGCAAGTTTCT AATTTCGGTTCAATAT CGATAGAGGAAAGT GTCT
cwp84	2054|2055a	AAAAAAGCTTATAA TTATCCTTAGTAGC CGTACCAGTGCGCC CAGATAGGGTG	CAGATTGTACAAATG TGGTGATAACAGAT AAGTCGTACCACCTA ACTTACCTTTCTTTGT	TGAACGCAAGTTTCT AATTTCGATTGCTACT CGATAGAGGAAAGT GTCT
***tcdA* (Toxin A)**	979|980a	AAAAAAGCTTATAA TTATCCTTATATCAC AGTTTTGTGCGCCC AGATAGGGTG	CAGATTGTACAAATG TGGTGATAACAGAT AAGTCAGTTTTCTTA ACTTACCTTTCTTTGT	TGAACGCAAGTTTCT AATTTCGGTTTGATAT CGATAGAGGAAAGT GTCT

3.2.2.1.4 Analysis of cwp84 mRNA

Total RNA was extracted as per General Methods 2.1.12.1 from stationary phase (24 hr) *C. difficile* 630ΔErm and CDΔCwp84$_{347}$. To isolate mRNA and remove residual genomic DNA contamination, the total RNA was treated with mRNA-ONLY™ Prokaryotic mRNA Isolation Kit (Epicentre Biotech) then TURBO DNA- free ™ Kit (Ambion). Reverse transcription PCR (RT-PCR), using primers amplifying *cwp84* either pre-intron insertion site (*cwp84*PreIntron_F CCTGTCTCAGCAGAAAACCA &

*cwp84*PreIntron_R GACCAA GATCCATTGGGTTG) or flanking the intron insertion site (*cwp84*-899-2334_F TCTCCTGAAGTAGCTTGTCC & *cwp84*-899-2334_R CCTGTCTCAGCAGAAAACC), was performed using SuperScript III One-Step RT-PCR system with Platinum® Taq DNA (Invitrogen).

3.2.2.2 Characterisation of *cwp84* knockout mutants

3.2.2.2.1 Growth rate determination

Crude spore suspension (30 µl) (produced as detailed in 3.2.2.2.2) was spread onto pre-reduced FAA plate(s) and incubated at 37 °C for 3 days in anaerobic conditions. sBHI broth (80 ml) was inoculated with 1 streak of colonies and incubated overnight as above. sBHI (80 ml) was inoculated with the resulting overnight culture, such that the OD_{600} ~0.1. The culture was then stirred and 1 ml samples removed at hourly intervals (including time 0) and OD_{600} measured. OD_{600} vs time was plotted and the linear (slope/gradient/exponential) part of the curve was analysed where appropriate.

3.2.2.2.2 Preparation of C. difficile spores

A portion of 48 hr *C. difficile* culture (300 µl) was spread onto FAA plates and incubated for 10 days. After incubation the bacterial lawn was scraped off and transferred to 15 ml of sterile DMEM and centrifuged for 20 min at 872 g. The resulting pellet was washed twice by resuspension into 12 ml DMEM followed by centrifugation and then resuspended into DMEM and heat shocked 62 °C for 40 min. The resulting crude spore suspension was aliquoted and stored at -80 °C. Samples (200 µl) were used to obtain viable counts after heat shocking.

3.2.2.2.3 Toxin A and B ELISA

Sandwich ELISA to toxins A and B was performed as detailed in General Methods Section 2.7.1.

3.2.2.2.4 Production of anti-Cwp84 peptides and antiserum production

The *cwp84* gene (*C. difficile* 630) was entered into MIF Bioinformatics epitope prediction program (September 2008). The following predicted antigenic peptides were synthesised and conjugated to keyhole limpet hemocyanin (KLH) by Peptide

Protein Research Ltd (Hampshire, UK) (numbers refer to Cwp84 amino acid position):

124	(C)TLEAYLKL	131
258	(C)APLNHAVAIV	267
513	(C)YETAVKVSQ	521

Three guinea pigs, one for each peptide, were injected with 800 µg peptide diluted in PBS and emulsified with Titremax™ adjuvant prior to injection into 2 sites subcutaneously. Boosts of 400 µg were administered at 21 and 42 days later. Animals were bled at 6 weeks and serum collected. Antibodies were purified from serum using a HiTrap Protein A column (General Methods section 2.4.1.2). Western blots were performed against low pH glycine extracts and cell wall extracts (Jonquieres *et al.* 1999; Wright *et al.* 2005) using guinea pig anti-Cwp84 peptide IgG at the following concentrations: 1:100, 1000, 5000, 7000, 1:10,000. Anti-guinea pig HRP (Sigma) was used at a 1:2000 dilution in blocking buffer. The following blocking reagents were tested, 5% milk, 3% BSA, protein-free blocking reagent (Pierce).

3.2.2.2.5 Cleavage of CDΔCwp84 SLPs with trypsin and other proteases

CDΔCwp84 SLPs, or CDΔCwp84 culture supernatant, was mixed with 1, 5 or 10 µg/ml trypsin (in 50 mM HEPES, 0.15 M NaCl pH 7.4) and incubated at room temperature. Samples were removed at time intervals and the reaction stopped by addition of trypsin inhibitor (final concentration 5 times that trypsin concentration). Samples were then analysed by SDS-PAGE.

CDΔCwp84 SLPs were also incubated with pre-activated clostripain (5 µg/ml) chymotrypsin (5 µg/ml), papain (1 or 5 µg/ml) and cathepsin B (50 µg/ml) and incubated at either room temperature or 37 °C. Samples were removed at time intervals and the reaction stopped by mixing with 4 x LDS loading buffer and heating at 95 °C for 5 min.

3.2.2.2.6 CDΔCwp84 grown in trypsin

CDΔCwp84 was grown in sBHI with addition of 0.1, 10 or 100 µg/ml trypsin for 24 hr. Cells were then harvested and washed once in 10 mM HEPES, 0.15 M NaCl pH 7.4 containing 500 µg/ml trypsin inhibitor. SLPs were then extracted, as

aforementioned (General Methods section 2.7.2), but using 0.2 M glycine pH 2.2 containing 100 µg/ml trypsin inhibitor.

3.2.2.2.7 Animal Model for Clostridium difficile Infection

Syrian hamsters (80-100 g) were housed (2 per cage) in isolator cages fitted with air filters on lids to minimise contamination between groups. Groups of 14 hamsters were divided into a 'Test' subgroup of 10 and 'Control' subgroup of 4 animals. All 14 hamsters were weighed and administered clindamycin (2 mg in 0.2 ml sterile H_2O) by the orogastric route on Day 0. On Day 2, hamsters in the 'Test' subgroup were challenged with 200 colony forming units of *C. difficile* spores in 0.2 ml DMEM, given orogastically. Control subgroup animals received just DMEM. All animals were weighed daily and monitored 6 times/day for 12 days for disease symptoms which included: diarrhoea, weight loss, lethargy and tender abdomen (Sambol *et al.* 2001; Babcock *et al.* 2006). Hamsters were scored on a 0-3 scale for disease symptoms and animals in advanced stages of disease or become immobile were euthanized. Faecal samples of euthanized animals were taken by colectomy and *C. difficile* isolated/cultured using alcohol shock method (HPA 2008) and plated onto *C. difficile* selective medium (E&O Labs) and grown for up to 7 days.

3.2.2.2.8 Microscopy of CDΔCwp84 mutant

Transmission electron microscopy (TEM) and Cryo-Scanning Electron Microscopy (Cryo-SEM) of whole cells was performed by Dr. H. Tolley (Health Protection Agency). Atomic force microscopy was performed by Dr. J. Mitchells at the Microscopy and Analysis Suite (MAS), University of Bath.

3.2.2.2.8.1 Transmission electron microscopy (TEM) – Whole cells

Cultures of WT and CDΔ*cwp84* were grown for 24 hr before harvesting and washing in PBS. Pellets were then resuspended with formalin and incubated for 16 hr before brief sonication. *C. difficile* cell suspension (2-3 µl) was applied to carbon film coated copper grids and incubated for 30 sec; surplus was removed with moist filter paper. Grids were stained with 2-3 µl 2% methylamine tungstate (pH 6.5) for 1-10 seconds before surplus removal as aforementioned. Grids were allowed to air dry

before examination on a Philips CM100 operating at 80kV (LaB_6 (lanthanum hexaboride crystal) filament).

3.2.2.2.8.2 Embedded sectioning TEM – Whole cells

WT and CDΔ*cwp84* 24 hr culture was harvested by centrifugation then resuspended in 2% glutaraldehyde buffered with 150 mM Sorensen's buffer pH 7.4 for 2 hr at 22 °C. Cells were pelleted by centrifugation then secondary fixed by re-suspension in osmium tetroxide for 3 hr at 4˚C. Cells were pelleted and gently re-suspended in molten 2% agarose, allowed to cool then cut in to 1 mm^3 cubes. Cubes were dehydrated through an ethanol gradient (30, 50, 70, 90, 100% for 10, 10, 20, 30 and 2 x 30 min respectively) followed by two incubations in 100% propylene oxide for 2 x 10 min. Cubes were then transferred to a 50:50 mixture of propylene oxide:Araldite™ CY212 resin mixture and incubated for first 2 hr, then overnight in fresh Araldite™ CY212 resin only at 22 °C. The cubes were transferred to embedding capsules and covered with fresh Araldite™ CY212 resin and incubated for 24 hr at 22 °C, followed by incubation at 60 °C for 72 hr to polymerise resin. Ultrathin 70 nm sections were prepared on a Reichert Ultracut E Ultramicrotome and transferred on to 400 mesh copper grids, followed by double staining with uranyl acetate then by lead citrate. Grids were carbon coated and examined using a Philips CM100 TEM as above.

3.2.2.2.8.3 Cryo-Scanning electron microscopy (Cryo-SEM)

C. difficile was spread onto FAA and grown for 24 hr (WT only) and 48 hr (WT and CDΔ*cwp84)* in anaerobic conditions. Bacterial colonies were carefully scraped from the plate and applied to a rivet. A second rivet was placed on top of the meniscus of bacterial culture. The rivet-bacteria-rivet assembly was snap frozen in liquid nitrogen slush, placed in a specimen holder (under liquid nitrogen) and transferred (frozen) into the Gatan Alto 2500 cryo-preparation unit. The top rivet was snapped off at a temperature of -184˚C. The specimen holder was warmed up to -95˚C and held at this temperature for 2 minutes to allow for sublimation of water from the fracture plane and then cooled down to below -140˚C prior to sputter coating with gold for 120 seconds (approximately 10 nm thickness of gold). The prepared

specimen was then transferred into the Philips XL30 FEG for SEM. All temperatures are those indicated on the Gatan Alto 2500 control pad and are not absolute values.

3.2.2.2.8.4 Atomic Force Microscopy (AFM)

3.2.2.2.8.4.1 Sample preparation

The native S-layer from *C. difficile* was extracted from 24 hr cultures of WT and CDΔCwp84$_{347}$ using low pH extraction (see General Methods 2.7.2) or using 50 mM tris 8 M urea pH 8.0 and incubating for 40 min at 22 °C, followed by centrifugation and 0.2 µm filtering. Samples were stored until processing at 4 °C.

3.2.2.2.8.4.2 Atomic Force Microscopy

Samples were dialysed for 12 hr using 1 kDa MWCO dialysis tubing against 15 mM PIPES pH 6.76 with 50 mM calcium chloride. The following surfaces were then immersed in the S-layer solution for 16 hr: graphite (positive surface charge - hydrophobic), ozone modified graphite (negative surface charge - hydrophilic) or freshly cleaved mica (negative surface charge – hydrophilic). Samples were removed from S-layer solution and washed with 10 mM potassium orthophosphate adjusted tris buffer pH 7.8 (with additional 50 mM calcium chloride for CDΔCwp84 S-layer imaging).

Atomic force spectroscopy was performed with a Nanoscope IIIa AFM (Brooker, Santa Barbara, USA) equipped with a J scanner (80 µm) and an oxide sharpened Si_3N_4 stylus on cantilevers with a spring constant of approximately 0.05 N/m (MSNL probes, Brooker, UK) with a radius of curvature ≤5 nm. Topographs were recorded in tapping mode (WT S-layer) or contact mode (CDΔCwp84 S-layer) at an approximate stylus loading force of 100 pN and a line frequency of typically 1Hz. AFM was performed at 22 °C in orthophosphate adjusted tris buffer pH 7.8 (with additional 50 mM calcium chloride for CDΔCwp84 S-layer imaging) using the fluid cell. AFM images of S-layers were basically processed by using the open-source Gwyddion AFM image processing tool (www.gwyddion.net).

3.3 Results

3.3.1 Production of *C. difficile cwp84* Knockout mutants.

Using the Sigma TargeTron algorithm, a total of 28 potential intron insertion sites were found in the *cwp84* gene. Out of the 28 the following sites were chosen (numbers represent base pairs from the start codon) (Table 3.3.1).

Table 3.3.1 Selected *cwp84* gene knockout target sites chosen from TargeTron algorithm

Insertion location	Gene 5`exon -intron-exon Gene 3'	Score	E-value	Functional insertion region
2054|2055a	TCTAAATTTGATGATAAAGTAGCTGTAC CA - intron - CCAATTATATAAGAC	8.73	0.063	Cell wall binding repeat domain
677|678a	CCTGATGTCACTGAACCATATTGCATTA TA - intron - GCATTTTTTACAGTT	7.54	0.153	Cysteine protease domain*
347|348a	TCTAAAGTTGACATACCTGAAAAAGAC CAG - intron - CATGTATTAAGACTT	7.30	0.183	Cysteine protease domain*
68|69a	TGGTTTTCTGCTGAGACAGGTATTGTTG AC - intron - ACTATTAAAAAGCAA	6.93	0.242	Signal Peptide
***Active site residues for Cwp84 are predicted to be at Gln-110 (328 bp), Cys-116 (346 bp), His-262 (784 bp), Asn-287 (859 bp) (Savariau-Lacomme *et al.* 2003).**				

Initially, insertion site 347|348a was chosen for investigation. However, a further three KO sites were selected to verify the outcome of the initial knockout and to determine whether the insertion location affected knockout outcome. The (353 bp) intron retargeting region for *cwp84* insertion sites 68|69a, 677|678a, 2054|2055a were produced by *de novo* gene synthesis, whereas 347|348a was retargeted by site SOE PCR.

Confirmation of *cwp84* knockout, i.e. intron integration, was performed by PCR using primers flanking *cwp84* (Figure 3.3.1 A). Confirmatory PCR of the Erm-RAM region also confirmed the Erm resistant phenotype was due to self-splicing of the

group I intron within the group II intron rather than a spontaneous erythromycin mutant (Figure 3.3.1 B).

RT-PCR of CDΔCwp84$_{347}$ mRNA using primers flanking the intron insertion site showed expression of the full length transcript in stationary phase cells from wild type *C. difficile* 630ΔErm and a complete absence of a similar size product from CDΔCwp84$_{347}$ (Figure 3.3.1 C(i)). However, primers amplifying *cwp84* mRNA pre-intron insertion site e.g. <345 bp, gave bands in both WT and CDΔCwp84$_{347}$. As expected primers amplifying *cwp84* post-intron insertion site, e.g. >348 bp, only gave a product for WT transcript, demonstrating truncation of the CDΔCwp84$_{347}$ *cwp84* transcript (Figure 3.3.1 C(ii)). It follows therefore, that the transcript of CDΔCwp84$_{2054}$ contains coding information for the cysteine protease domain and one of the three CWB domains.

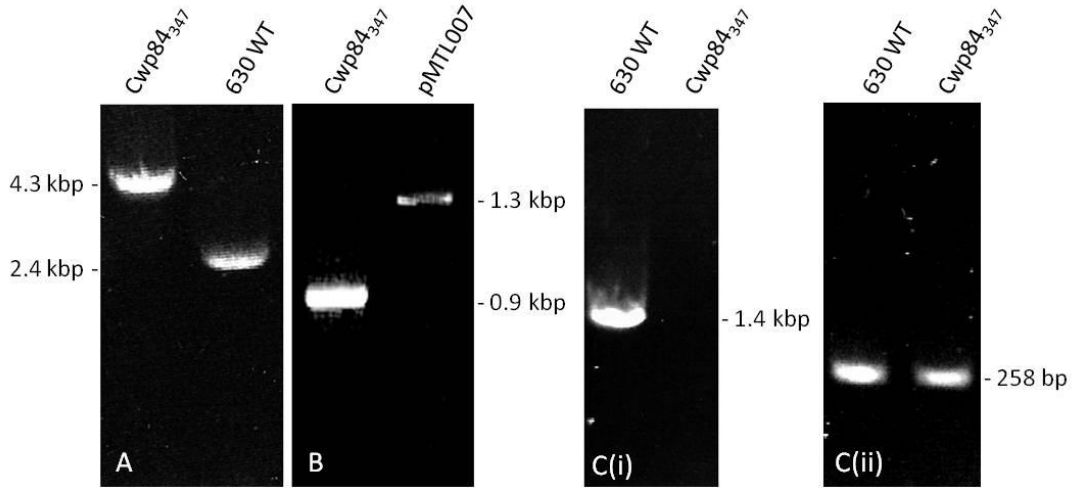


Figure 3.3.1 Genetic Characterisation of Cwp84 Mutant. (A) Confirmation of intron integration in *cwp84* gene by PCR. Lane 1, 4.3 kbp product from CDΔCwp84$_{347}$. Lane 2, 2.4 kbp product from *C. difficile* 630ΔErm. (B) Demonstration of Erythromycin Retrotransposition-Activated Marker (Erm-RAM) self-splicing, an event strictly coupled to group II intron integration. Lane 1, 900 bp product from CDΔCwp84$_{347}$. Lane 2, 1300 bp product from native pMTL007-EC2 shuttle vector. (C) Reverse transcription PCR (RT-PCR) of *cwp84* mRNA (C(i)) flanking intron integration site in stationary phase cultures. (C(ii)) Pre-intron insertion site. Lane 1, *C. difficile* 630ΔErm; Lane 2 CDΔCwp84$_{347}$.

3.3.2 Assessment of other genes potentially affected by intron insertion

To confirm that any phenotypic effects observed were a direct result of *cwp84* KO and not due to intron promiscuity, analysis of other potentially affected genes was undertaken.

Analysis of the *cd1751* gene was undertaken since this shows significant identity (70.4%) to the *cwp84* gene and encodes another putative *C. difficile* surface-associated cysteine protease in *C. difficile* 630. The *cd1751* gene is also highly conserved with 99.0% identity between 630 (ribotype 012) and R20291 (ribotype 027). RT-PCR of WT and CDΔcwp84$_{347}$ KO mRNA using primers flanking *cd1751* does not give a product, suggesting *cd1751* is not expressed in stationary phase. PCR analysis of *cd1751* in all *cwp84* KO mutants showed this gene to be unmodified by the gene knockout procedure. Furthermore, erythromycin resistant (indicating possible intron integration) but *cwp84* KO negative mutants were checked for integration into *cd1751*. All screened *cwp84* KO negative clones had WT (2.4 kbp) *cd1751* indicating either insertion elsewhere in the genome or were spontaneous erythromycin mutant(s).

Due to the observed effects on SlpA maturation, *slpA* was sequenced from CDΔCwp84 mutants to ensure that the resultant phenotype was due to *cwp84* KO rather than mutations, introduced either accidentally or in *slpA* as a result of *C. difficile* survival mechanisms, which caused the observed changes in SlpA processing. No mutations were found suggesting that the effect on cleavage was a result of *cwp84* KO and specific to *cwp84* inactivation. Intron insertion into *slpA* was also assessed by PCR, no insertion was found in *slpA* in all created *cwp84* mutants (Table 3.31). Intron insertion into the 26 other CWBD containing protein genes were not assessed.

3.3.3 Morphology and growth characteristics of cwp84 KO mutants

3.3.3.1 Morphological

Cwp84 KO mutants grown on solid medium showed significantly different colony morphology compared to wildtype *C. difficile* 630ΔErm. While the latter showed a characteristic irregular, translucent appearance, CDΔCwp84 formed more regular circular colonies which were creamy white in appearance (Figure 3.3.2).

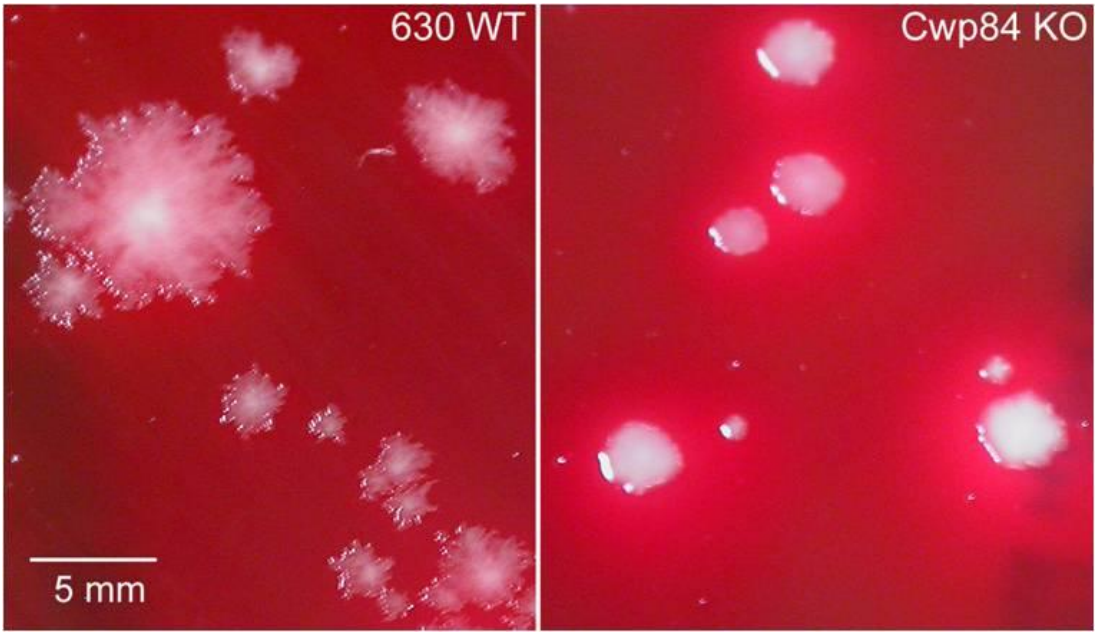


Figure 3.3.2 Difference in colony morphology of wildtype *C. difficile* 630ΔErm and CDΔCwp84$_{347}$ grown on FAA plus 5% horse blood.

Gram stain of CDΔCwp84 cells from solid phase 24 hr and 48 hr cultures appeared indistinguishable from the wildtype *C. difficile* 630ΔErm, however 48 hr CDΔCwp84 cultures appeared largely as pink 'Gram negative' cell debris.

3.3.3.2 Growth

In liquid culture CDΔCwp84 grew slower than wildtype *C. difficile* 630ΔErm (Table 1.3.2) and at higher culture ODs CDΔCwp84 appeared to be unable to maintain a prolonged stationary phase and proceeded rapidly into death phase (Figure 3.3.3). CDΔCwp84 post death phase cultures maintain a culture OD_{600} ~0.4, a value reached and maintained by the WT after 48 hr post inoculation.

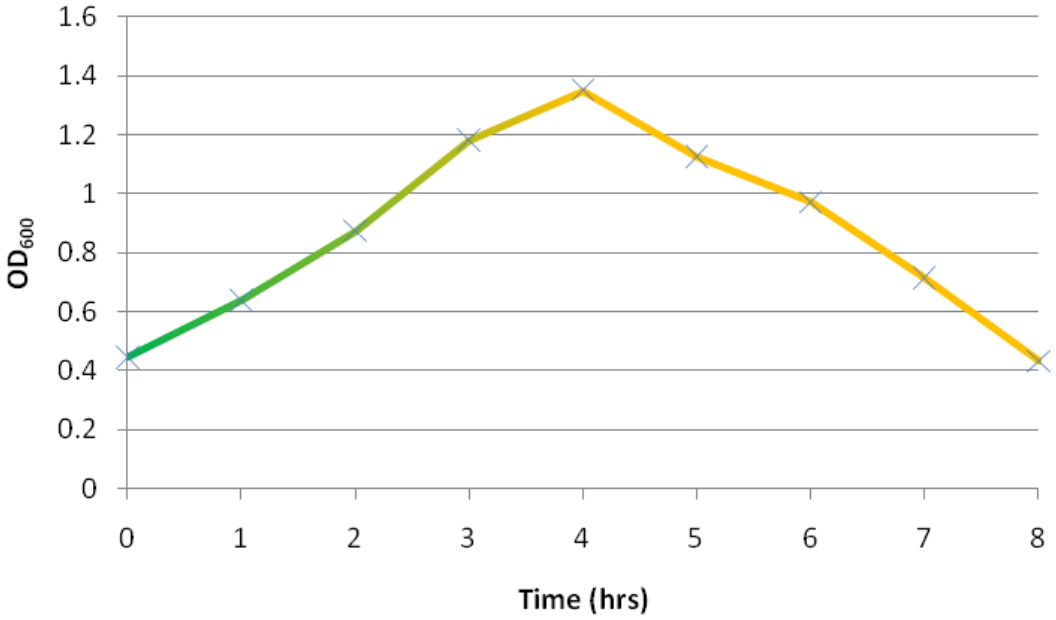


Figure 3.3.3 Decline in culture OD_{600} of CDΔCwp84$_{347}$ over time. An entire plate culture (24hr) was scraped off and re-suspended into culture media (time 0) and OD_{600} measured at hourly intervals during growth. Note: plate cultures were used to inoculate, to prevent dilution of any potential growth altering substance found within liquid starter cultures.

In line with the decrease in culture OD during growth of CDΔCwp84, is an increase in culture viscosity and propensity of the culture to aggregate and form stringy precipitates, particularly on the walls of a stirred culture vessel.

Growth rates of a *tcdA* (Toxin A) knockout were similar to the WT (Table 3.3.2) suggesting slower growth was due to *cwp84* KO and not a general consequence of the knockout process.

3.3.3.3 Sporulation

Cwp84 mutants retained their ability to sporulate, albeit at reduced efficiency compared to the wildtype. The level of sporulation, estimated from viable counts before and after heat shock treatment was determined to be 0.64 ± 0.03% (64.19 ±0.14% Log cfu) for 18 day cultures of CDΔCwp84$_{347}$ compared to 7.18 ± 3.45% (78.41 ±6.11% log cfu) for wildtype *C. difficile* 630ΔErm.

Table 3.3.2 Growth rates of *C. difficile* mutants

***C. difficile* Strain**	**Growth Rate (OD_{600} h^{-1})**	**%WT**
C. difficile 630ΔErm	0.29 ± 0.011	100%
CDΔCwp84$_{347}$	0.12 ± 0.007	42%
CDΔToxinA$_{980}$	0.26 ± 0.018	90%
C. difficile 630ΔErm + trypsin*	0.30 ± 0.002	103%
CDΔCwp84$_{347}$ + trypsin (µg/ml)		
0	0.12 ± 0.002	42%
1	0.13 ± 0.008	43%
3	0.16 ± 0.003	55%
10	0.17 ± 0.003	58%
100	0.21 ± 0.018	71%

Growth rates were measured on the linear portion of the growth curve (n=3)

*Trypsin added to growth medium at100 µg/ml

3.3.3.4 Toxin production of *cwp84* mutants

Toxins A and B are produced when cultures reach stationary phase (usually after 15 hr, 24 hr for *C. difficile* 630 (Merrigan *et al.* 2010)) and when cultures are stressed, typically though nutrition depletion. The ability of CDΔ*cwp84*$_{347}$ to produce toxin was assessed by ELISA and compared to toxin production of the wild-type under similar growth conditions.

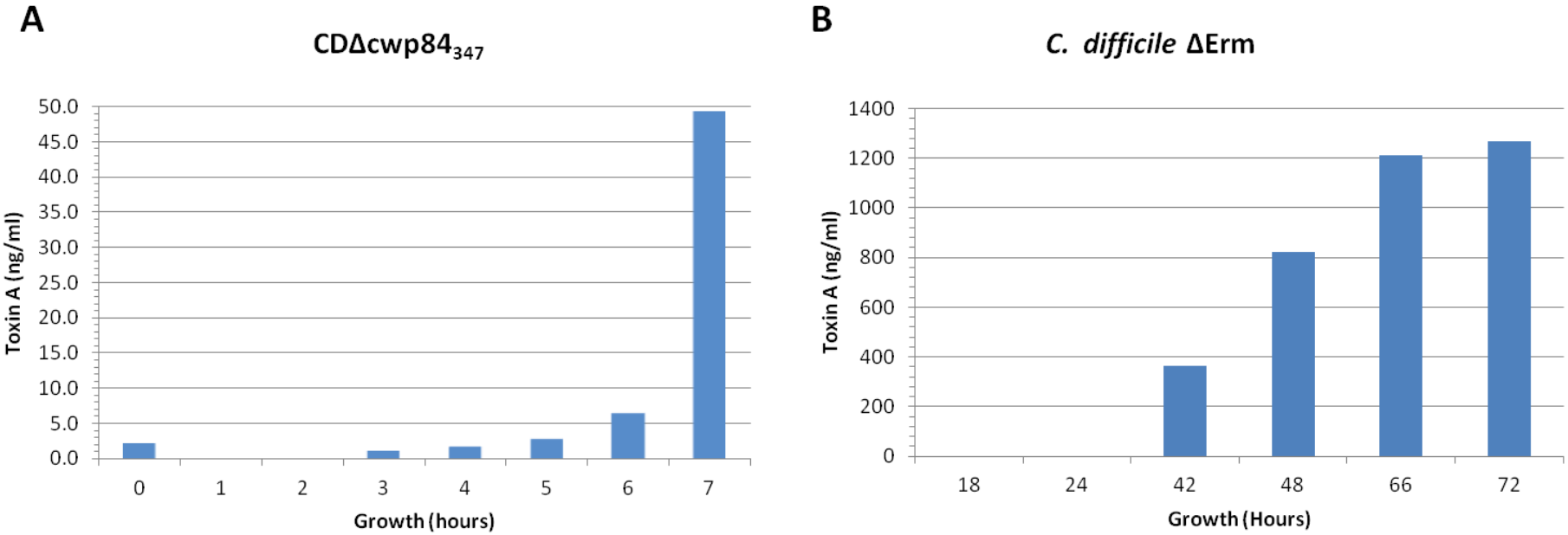


Figure 3.3.4 Toxin A ELISA of culture supernatant fluids from CDΔCwp84$_{347}$ and wildtype *C. difficile* 630ΔErm grown in sBHI. (A) CDΔCwp84$_{347}$, taken at hourly intervals (culture inoculated from solid phase media). (B) Wildtype *C. difficile* 630ΔErm over 3 days at 6 and 18 hr intervals. Limit of Quantification: 5 ng/ml.

Culture supernatant fluids of CDΔCwp84$_{347}$ contained approximately 50 ng/ml toxin A after 7 hr (Figure 3.3.4 A) and after 24 hr contained ~ 860 ng/ml toxin A, whereas no quantifiable toxin A was found in the culture supernatant of *C. difficile* 630ΔErm until *after* 24 hr (Figure 3.3.4 B). Culture supernatants from 24 hr growth of a *tcdR(D)* mutant did not contain any toxin, whereas 24 hr culture supernatant from a Spo0A mutant contained approximately 260 ng/ml toxin A (Chapter 5 Section 5.3.3.2). The culture supernatant of other surface protein mutants also demonstrated higher toxin A levels compared to the WT after 24 hr, particularly a FliD (flagella cap) KO (Chapter 5 Section 5.3.3.2). This data suggests that higher culture supernatant toxin A levels are not specific to CDΔCwp84$_{347}$ and may be associated with changes in toxin gene expression and/or a decrease in cellular integrity i.e. intracellular toxin is leaking out more easily.

3.3.3.5 Anecdotal results

Analysis of the culture supernatant from CDΔCwp84 grown in incompletely reduced media showed a large number of protein species (other than those normally visible without concentration) suggesting a large degree of cell lysis. Furthermore, CDΔCwp84 stored in poorly sealed anaerobic jars were not viable. In contrast, WT growth was unaffected by either incompletely reduced media or microaerobic environments. These two anecdotal results suggest that the Cwp84 mutants appear to be more sensitive to oxygen levels than the WT.

C. difficile is known to produce a characteristic smell, primarily due to production of volatile organic compounds (VOCs) including *p*-cresol (Dawson *et al.* 2008). The smell of the *cwp84* mutant is very much stronger and sharper on the nose. Examination of the pattern of VOCs on gas-liquid chromatography may elucidate individual components or fermentation pathways creating this unique smell, particularly those differing from the WT.

3.3.4 Analysis of surface layer proteins of cwp84 mutants

Previous studies have demonstrated that a large number of the SLPs of *C. difficile* can be conveniently extracted by several methods including low pH (Calabi *et al.* 2001; Wright *et al.* 2005).

SDS-PAGE of SLPs extracted from wildtype *C. difficile* 630ΔErm by this method showed a characteristic band pattern in which the two principal SLPs, derived from the post translational cleavage of the single precursor, SlpA, appeared as bands of approximately 39 kDa and 48 kDa corresponding to the LMW SLP and HMW SLP, respectively (Figure 3.3.5A).

In contrast, these bands were absent in SLP extracts of CDΔ*cwp84* in which a new prominent band of approximately 84 kDa was evident by SDS-PAGE analysis. This band was subject to proteomic analysis (GeLC-MS/MS) using in-gel tryptic digestion. Mascot individual ion scores of >47 for 13 individual peptides positively identified the 84 kDa band as the SlpA precursor protein (predicted molecular mass of 76 kDa). N-terminal sequence analysis of the SlpA precursor protein band gave the amino acid sequence ATTGT which is identical to that of the N-terminus of mature LMW SLP and demonstrated that the signal sequence has been removed from the precursor protein. These data show that in the CDΔCwp84 mutant, maturation of the SlpA precursor protein is incomplete, in that following removal of the signal peptide, cleavage into the HMW and LMW SLPs fails to occur.

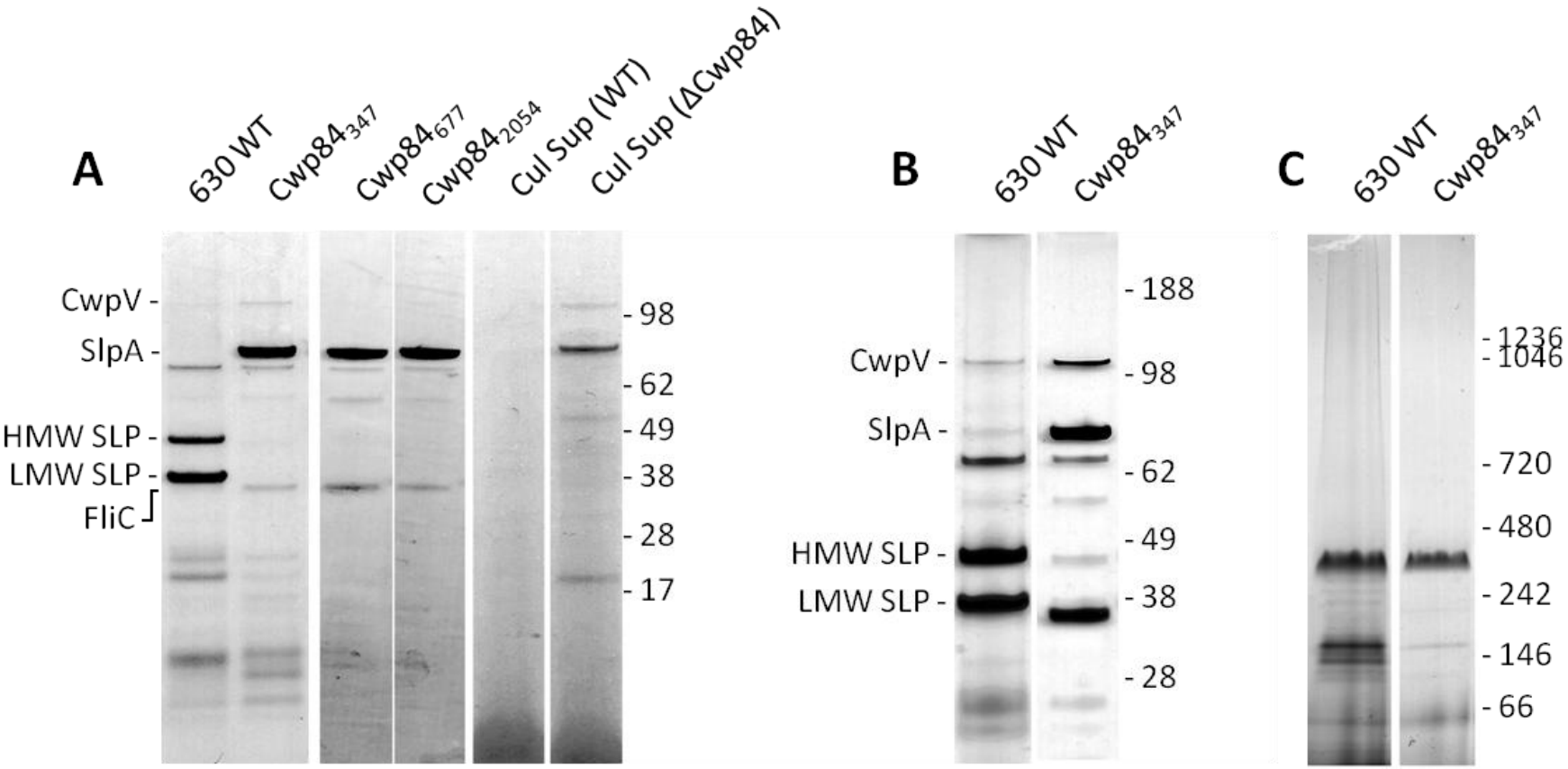


Figure 3.3.5 Comparision **of** *C. difficile* 630ΔErm and CDΔCwp84 surface protein extracts. (A) Analysis of extracted SLPs from wildtype *C. difficile* 630ΔErm, CDΔCwp84$_{347}$, CDΔCwp84$_{677}$ and CDΔCwp84$_{2054}$. Lanes 5 and 6, 24 hr culture supernatant fluids from *C. difficile* 630ΔErm (Cul sup WT), CDΔCwp84$_{347}$ (Cul sup ΔCwp84), respectively. (B) Analysis of extracted SLPs by silver stained SDS-PAGE from wildtype *C. difficile* 630ΔErm and CDΔCwp84$_{347}$. (C) Native PAGE (BN PAGE) analysis of extracted SLPs wildtype *C. difficile* 630ΔErm and CDΔCwp84$_{347}$ (concentrations normalised).

Another characteristic of CDΔCwp84 is that the SlpA precursor protein was evident in the culture supernatant of 24 hr cultures (Figure 3.3.5A), unlike wildtype *C. difficile* 630ΔErm where there was no evidence of any SLPs being released under similar growth conditions. Brisk agitation of the CDΔCwp84 and WT cultures did not appear to increase the concentration of SlpA in the culture medium nor release any SLPs from the WT, suggesting the loss of SlpA is not a consequence of physical sheering.

Blue native PAGE analysis reveals that while SlpA of CDΔCwp84$_{347}$ is not matured into its component SLPs, it can also form a 320 kDa species similar to WT SLPs (Figure 3.3.5C), as originally demonstrated by size exclusion chromatography (Fagan *et al.* 2009).

3.3.4.1 Identification of other surface proteins in *cwp84* KO SLPs

The *cwpV* gene (*cd0514*) encodes a predicted surface protein of 167 kDa. Recently, it has been shown by analysis of surface protein extracts that CwpV is processed into two fragments: a 40 kDa N-terminal cell wall binding fragment and a 120 kDa C-terminal fragment (Emerson *et al.* 2009b). Silver stained SDS-PAGE of SLPs extracted from wildtype *C. difficile* 630ΔErm and CDΔCwp84$_{347}$ both showed the presence of a band of approximately 115 kDa (Fig 3.3.5B). Proteomic analysis of the 115 kDa band from CDΔCwp84$_{347}$ identified this band as the 120 kDa C-terminal fragment of CwpV. These data suggest that the maturation of CwpV is not affected by knockout of *cwp84*. Moreover, a protein of equivalent size (115 kDa) is found in the culture supernatant of CDΔCwp84, putatively indicating that CwpV is also lost from the cell surface similar to immature SlpA.

Silver stained SLP extracts from both WT and CDΔCwp84$_{347}$ demonstrate how, other than the loss of the main S-layer bands, the SLP extracts appear similar. Bands at 84, 66, 58, 45 and 36 kDa are present in both extracts. The presence of an 84 kDa band (identified as SlpA in CDΔCwp84$_{347}$ SLPs) in WT SLPs (and in R20291, VPI10463 and ARL001 SLPs, Chapter 4 Section 4.3.1) may suggest that the SLP extraction may have also removed immature SlpA, presumably that which has yet to be cleaved.

Due to its prominence in CDΔCwp84 SLPs (Figure 3.3.5A), the 36 kDa band was subjected to N-terminal sequence analysis and gave the amino acid sequence MRVNT. BLAST searching this sequence against the *C. difficile* 630 proteome, assuming no post-translational modifications, matches the first 5 amino acids of FliC the flagella core/subunit protein.

Freeze thaw cycles of CDΔCwp84$_{347}$ SLPs result in the appearance of a 45 kDa band (Figure 3.3.5B) while other bands remain unchanged, suggesting the 45 kDa band resulted from degradation of immature SlpA. No protease contamination or auto-processing/maturity ability was found (Section 3.3.6.3) confirming the likely sensitivity of SlpA to freeze-thaw cycles.

3.3.5 Identification of Cwp84 at the cell surface.

It is suggested that Cwp84 can be found in cytoplasmic (90 kDa), membrane, cell wall associated (82 kDa) and surface (low pH) fractions (primarily 85 kDa), as determined by immunoblotting with mCwp84 antibodies (Janoir *et al.* 2007). To confirm the loss of Cwp84 from the cell surface of CDΔ*cwp84*$_{347}$, two methods were used.

A proteomic approach was assessed, whereby GeLC-MS/MS was used to identify proteins found between a 62-100 kDa region in 1D SDS-PAGE of WT and *cwp84* KO SLPs. Mass spectrometric analysis did not reveal the presence of Cwp84 in WT (or a lack of in CDΔCwp84$_{347}$) SLPs between 62-100 kDa. A variety of metabolic proteins were identified in WT SLPs present together with CwpV, CD2791 and CD2784, peptidases PepV and PepD. Analysis of the CDΔCwp84$_{347}$ SLPs region resulted in 5 identified proteins – CwpV, PepV, SlpA, CD2791 and CD2784, no metabolic associated proteins were found (Table 3.3.3). Reasons for the lack of metabolic protein 'contamination' in the 62-100 kDa region CDΔCwp84$_{347}$ SLPs are unknown.

To identify Cwp84 by immunoblotting, antibodies to synthetic peptides to predicted epitopes of Cwp84; covering the cysteine protease domain (104-322 aa) (T124-L131

and A258-V267 aka CP1 and CP2), and the first CWB domain (507-595 aa) (Y513-Q521 aka CWB-1), were produced in guinea pigs.

Unfortunately, immunoblotting WT SLPs or cell wall extracts (by method of Jonquières *et al.* (1999)) with the guinea pig serum or purified IgG for each of the peptides or combined, did not detect any unique bands (or a lack of in CDΔ*cwp84*$_{347}$ SLPs). Optimisation of antibody concentrations, blocking material or chemiluminescent reagents did not change the result. Blots using the CWB-1 epitope serum/antibodies did detect some specific bands, however these were likely to be CWB repeat (PFam04122) containing proteins or fragments. Confidence in the obtained results was undermined by strong cross reactions with molecular weight markers for all antiserum.

Cwp84 was therefore not found in between 62-100 kDa in low pH SLPs using a proteomic approach or by immunoblotting SLPs or other protein extracts of *C. difficile* with antiserum raised to peptides of predicted epitopes of Cwp84.

Table 3.3.3 Proteins identified by GeLC-MS/MS between a 62-100 kDa region excised from a 1D SDS-PAGE gel of WT and *cwp84* KO low pH SLPs.

WT	KO
WT1 • gi|126698095|ref|YP_001086992.1|- **CD0514** (cwpV) • gi|126700400|ref|YP_001089297.1|- **CD2784 (Cwp6)** • 3-hydroxybutyryl-CoA dehydrogenase (organic acid metabolism) • Rubredoxin oxidoreductase (iron binding protein) • Enolase (phosphopyruvate dehydratase) • Acetyl-CoA acetyltransferase	KO1 • gi|126698095|ref|YP_001086992.1|- **CD0514** (cwpV) • gi|126700719|ref|YP_001089616.1|- PepV (CD3102) Peptidase
WT2 • gi|126700400|ref|YP_001089297.1|- **CD2784 (Cwp6)** • Acetyl-CoA acetyltransferase • gi|126697752|ref|YP_001086649.1|- NAD-specific glutamate dehydrogenase • Electron transfer flavoprotein alpha and beta subunits • Rubredoxin oxidoreductase (iron binding protein)	KO2 • gi|126700409|ref|YP_001089306.1| - **SlpA**
WT3 • gi|126700400|ref|YP_001089297.1|- **CD2784 (Cwp6)** • Acetyl-CoA acetyltransferase • Rubredoxin oxidoreductase (iron binding protein) • gi|126700719|ref|YP_001089616.1|- PepV (CD3102) Peptidase • gi|126700078|ref|YP_001088975.1| - DnaK (CD2461) - Molecular chaperone	KO3 • gi|126700409|ref|YP_001089306.1| - **SlpA**
WT4 • gi|126700407|ref|YP_001089304.1| - **CD2791 (Cwp2)** • gi|126700400|ref|YP_001089297.1|- **CD2784 (Cwp6)** • gi|126700078|ref|YP_001088975.1| - DnaK (CD2461) - Molecular chaperone • gi|126698287|ref|YP_001087184.1| - PepD (CD0708) - putative aminoacyl-histidine dipeptidase	KO 4 • gi|126700407|ref|YP_001089304.1| - **CD2791 (Cwp2)** • gi|126700409|ref|YP_001089306.1| - **SlpA** • gi|126700400|ref|YP_001089297.1|- **CD2784 (Cwp6)**

The 62-100 kDa region was divided into four sub-regions (WT1-4 or KO1-4, decreasing in MW) prior to analysis. Proteins in Blue are CWBD containing proteins.

3.3.6 Effect of proteases on SlpA *in vitro* and effect on Cwp84 mutants

3.3.6.1 Effect of trypsin treatment on the SlpA precursor *in vitro*

Previous studies have suggested that SlpA is matured into its component HMW and LMW parts by cleavage at the Ser345-Ala346 bond within the sequence: ETKSANDT (Eidhin *et al.* 2006). Since there is a lysine residue close to the natural cleavage site of the SlpA precursor, the ability of trypsin to generate analogs of the LMW and HMW SLPs was assessed.

Treatment of CDΔCwp84 SLPs with a low concentration of trypsin (1 µg/ml) resulted in the disappearance of the 84 kDa (SlpA) band and concomitant appearance of two bands of similar size to the HMW and LMW SLP bands from wildtype *C. difficile* 630ΔErm (Figure 3.3.6A).

GeLC-MS/MS analysis identified the 39 and 48 kDa bands as the LMW and HMW SLPs, respectively. In addition, N-terminal sequencing of the 48 kDa (HMW) band gave the sequence SANDT, consistent with both predicted native maturation site and the expected trypsin cleavage site. N-terminal sequencing of the 39 kDa (LMW) band gave the sequence ATTGT, consistent with the predicted N-terminus of the LMW SLP. Similar results were obtained when higher trypsin concentrations (5 and 10 µg/ml) were used.

The LMW SLP appeared relatively resistant to trypsin digestion, while the HMW SLP appeared to be slowly degraded (Figure 3.3.6A). Assessment of the culture supernatant 'shed SlpA' from CDΔCwp84 also demonstrated the same cleavage by trypsin to give bands migrating at identical mobility to the WT S-layer bands.

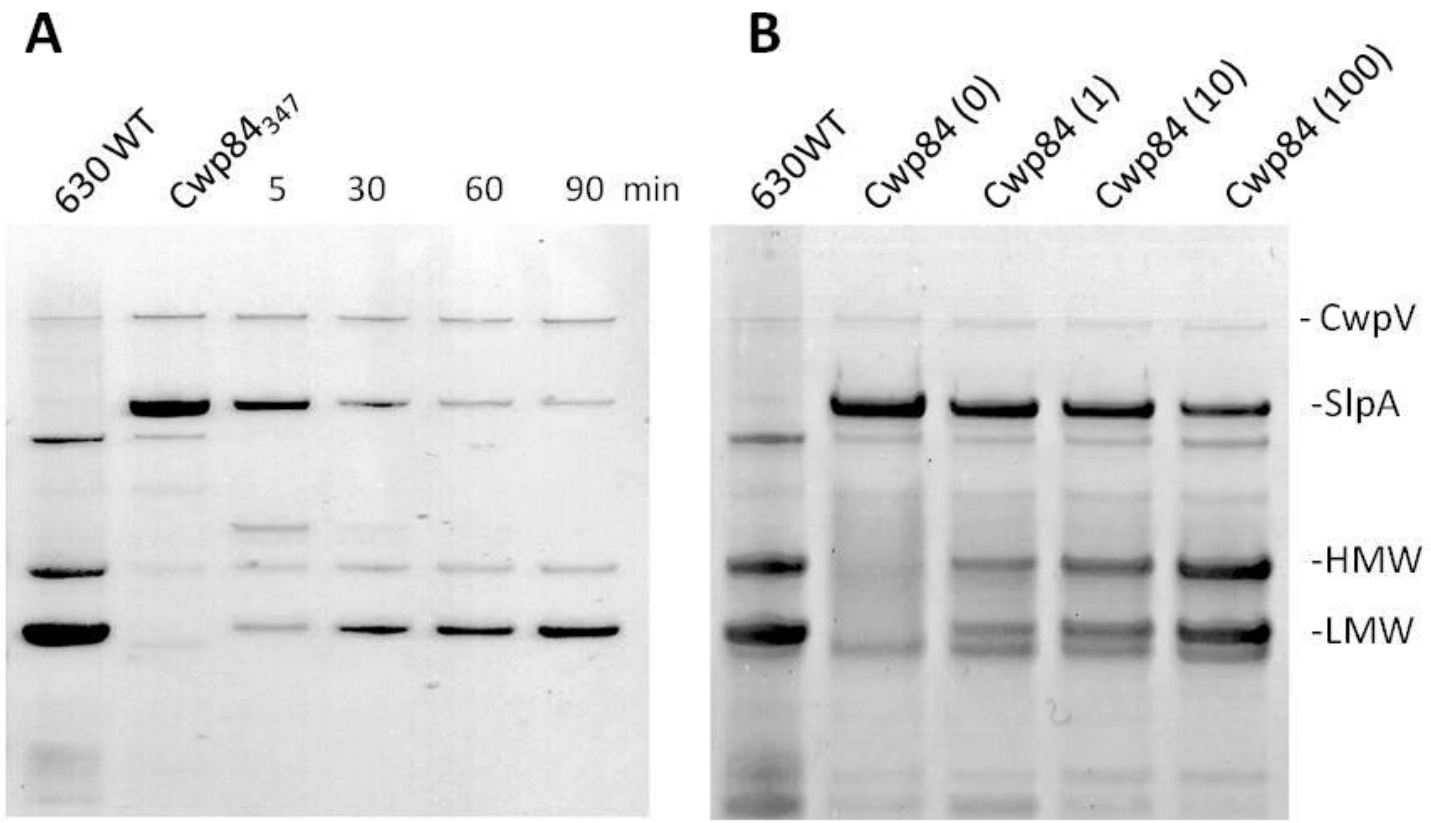


Figure 3.3.6 Effect of trypsin on the SLPs of CDΔCwp84. (A) Cleavage of SLP extracts from CDΔ*cwp84*$_{347}$ with 1 µg/ml trypsin (DPCC-treated, bovine pancreas; Sigma T1005) in 50 mM HEPES pH 7.4 containing 0.15 M NaCl at 22^{o}C for various times. (B) Extracted SLPs from 24 hr cultures of CDΔ*cwp84*$_{347}$ grown in sBHI supplemented with 0, 1, 10 and 100 µg/ml trypsin. Prior to low pH extraction of SLPs, cultures were washed 10 mM HEPES containing 100 mM NaCl pH 7.4 and 0.5 mg/ml trypsin inhibitor and then extracted in 0.2 M glycine pH 2.2 buffer containing 0.1 mg/ml trypsin inhibitor.

The 120 kDa band (CwpV) appeared to also be resistant to trypsin cleavage, while other components of the SLP extract tended to be degraded to completion (Figure 3.3.6).

Using trypsin inhibitor rather than heat inactivation, CDΔ*cwp84* SLPs incubated with 1 µg/ml trypsin appear to be cleaved into a multitude of fragments between 49-62 kDa. A WT-like 38 kDa LMW SLP band appears first, followed by a WT-like HMW SLP, then the degradation of other 49-62 kDa fragments (Figure 3.3.7).

3.3.6.2 Effect of trypsin on growth and SLPs of Cwp84 KO mutants

As the extracted SlpA precursor protein in CDΔCwp84$_{347}$ SLPs is rapidly cleaved in solution by trypsin, the ability of trypsin to cleave SlpA displayed on the surface of the bacterium was also assessed.

CDΔCwp84$_{347}$ was grown in medium supplemented with various concentrations of trypsin followed by washing with media containing trypsin inhibitor. Figure 3.3.6B shows SLPs extracted from cells treated in this manner compared to the SLP extracts from the untreated wildtype and CDΔCwp84$_{347}$ cultures. N-terminal sequences (SANDT and ATTGT) of the prominent bands which appeared as a result of trypsin treatment confirmed their identity as the HMW and LMW SLPs, respectively. The data shows that >50% of the SlpA precursor protein is cleaved on cells grown in the presence of the highest trypsin concentration used. Compared to trypsin treatment of the extracted SlpA, the HMW SLP appeared more resistant to trypsin degradation which may reflect its proposed orientation within the S-layer, in which the HMW SLP is the lesser exposed of the two subunits.

Addition of trypsin to the CDΔCwp84$_{347}$ growth medium significantly increased its growth rate in a dose dependent manner (Table 3.3.2) consistent with the cleavage of SlpA (Figure 3.3.6 B). The presence of trypsin in the growth medium also eliminated the propensity of CDΔCwp84 to aggregate; furthermore immature SlpA was not found in the culture supernatant in trypsin supplemented media.

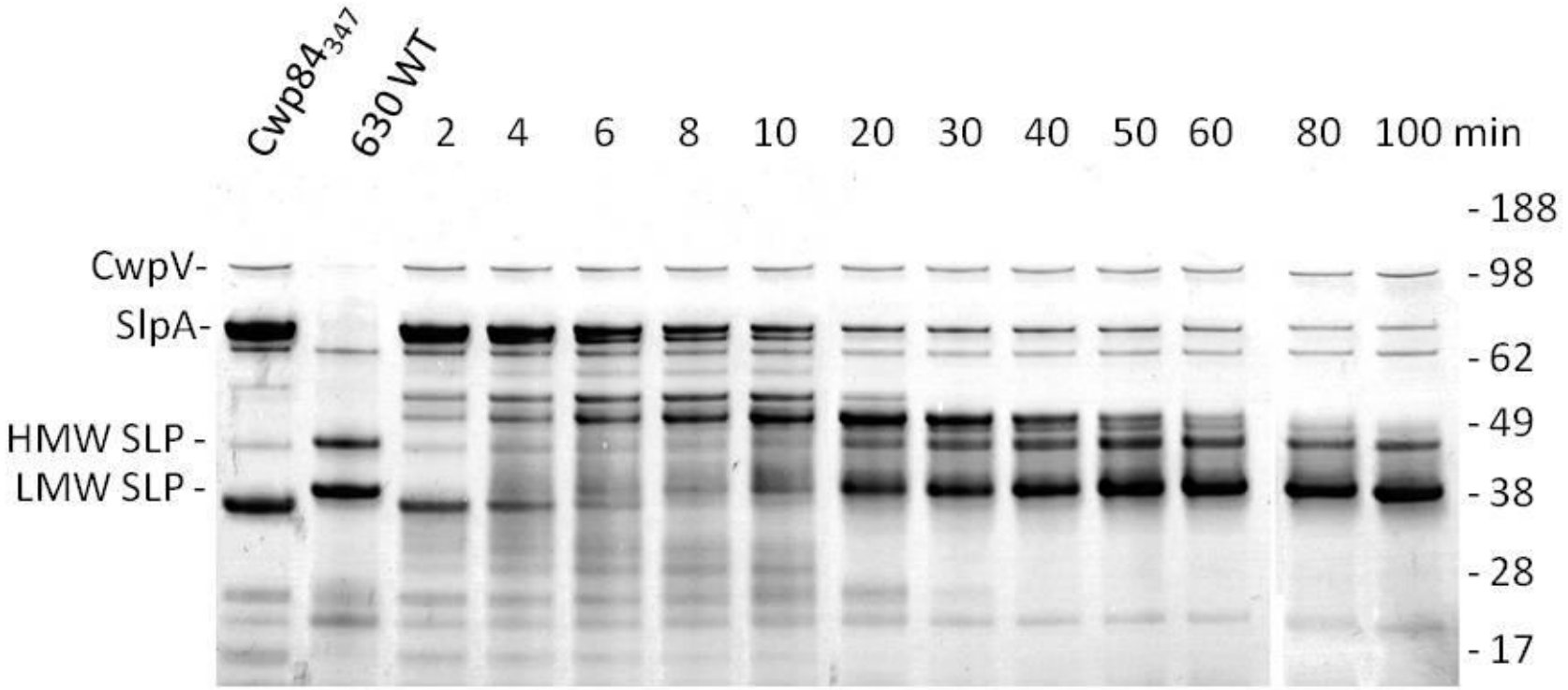


Figure 3.3.7 Termination of trypsin mediated cleavage of CDΔ*cwp84*$_{347}$ SLPs with trypsin inhibitor. Analysis of trypsin cleavage of SLP extracts from CDΔ*cwp84*$_{347}$ with 1 µg/ml trypsin in 50 mM HEPES pH 7.4 containing 0.15 M NaCl at 22^{o}C for various times. Reaction stopped by addition of 0.5 mg/ml trypsin inhibitor.

3.3.6.3 Effect of other proteases on the SlpA precursor

To test the hypothesis that SlpA is cleaved by Cwp84, investigations were made as to whether other members of the cysteine protease C1A family, to which Cwp84 putatively belongs, could cleave the SlpA precursor *in vitro* to give WT-like SLPs (Figure 3.3.8).

Treatment of CDΔCwp84 SLPs with a low concentration of papain (1 µg/ml) appeared to partially cleave SlpA giving bands at 57, 50, 48 and 38 kDa, while incubation with a higher concentration (10 µg/ml) appeared to digest SlpA leaving a 48 kDa band above highly fragmented lower MW species.

Treatment of CDΔCwp84 SLPs with cathepsin B required a relatively high (50 µg/ml) concentration before cleavage of SlpA was seen. Cathepsin B cleaved CDΔCwp84 SLPs into a highly fragmented band pattern between 84-47 kDa, with the original lower MW SLP extract bands (<38 kDa) remaining unaffected.

Two other digestive enzymes were also assessed for their ability to cleave SlpA. Chymotrypsin, at 5 µg/ml, appeared to partially digest CDΔCwp84 SLPs to give a band at 45 kDa, slightly lower than the WT HMW SLP (48 kDa), with a co-appearance of a 29 kDa band. The 74 kDa total of these two 45 and 29 kDa bands, although less than the observed 84 kDa for SlpA, correlates well with the predicted molecular weight of SlpA of 74 kDa, suggesting they are derived from the cleavage of SlpA rather than other SLP components. The ability of elastase to cleave the Ser345-Ala346 bond within SlpA was also assessed. Like trypsin, elastase appeared to cleave SlpA in CDΔCwp84 SLPs into two bands at 39 and 45 kDa, rather than 39 and 48 kDa for WT SLPs.

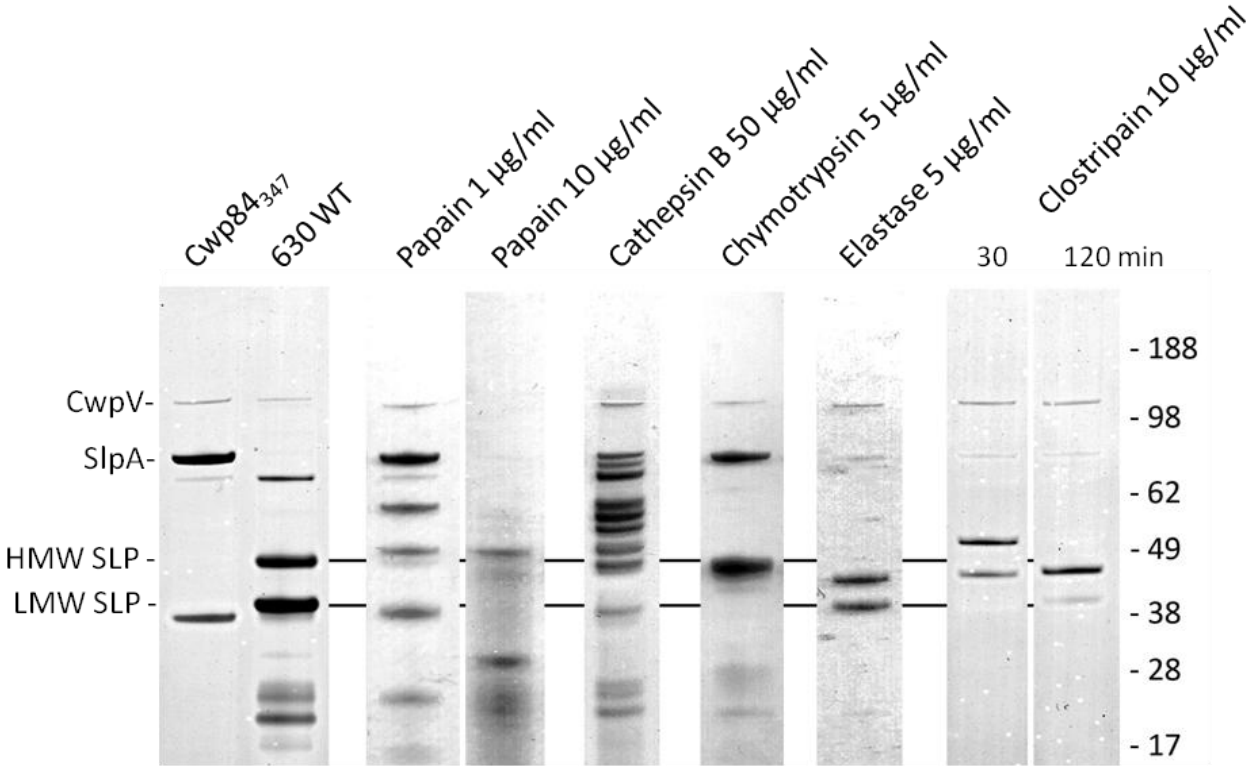


Figure 3.3.8 Effect of proteases on SlpA in CDΔ*cwp84*$_{347}$ SLPs. Extracts incubated with: Papain (Sigma 76220) at 1 and 10 µg/ml in 50 mM HEPES plus 0.15 M NaCl pH 7.4 at 22 °C, after 4 hr and 30 min respectively. Cathepsin B (Sigma C0150) at 50 µg/ml in 50 mM HEPES plus 0.15 M NaCl pH 6.0 at 40 °C, after 5 hr. Chymotrypsin (Sigma C3142) at 5 µg/ml in 50 mM HEPES plus 0.15 M NaCl pH 7.4 at 22 °C, after 4 hr. Elastase (Sigma E0258) at 5 µg/ml in 50 mM HEPES plus 0.15 M NaCl pH 7.4 with 50 µg/ml trypsin inhibitor at 22 °C, after 2 hr. Clostripain (Sigma C0888) at 5 µg/ml in 50 mM HEPES 10 mM $CaCl_2$ 2.5 mM DTT pH 7.6 at 22 °C (pre-activated for 3 hr), after 30 min and 2 hr. Lines across figure represent reference MW to HMW and LMW SLP bands.

The surface located thiol protease, eluded to by Seddon & Borriello (1992), was suggested to have trypsin-like characteristics, but an inhibition profile more akin with C11 family cysteine protease clostripain from *C. histolyticum*. Incubation of CDΔCwp84 SLPs with 10 µg/ml clostripain initially gave two bands at 47 and 51 kDa after 30 min, 40 and 47 kDa bands after 120 min.

It is interesting to note that 120 kDa fragment of CwpV was not cleaved by trypsin, cathepsin B, chymotrypsin, clostripain or elastase but was degraded by papain at 10 µg/ml.

To ensure that SlpA was not being cleaved by SLP extract protease contamination or via auto-maturation, CDΔCwp84 SLPs were incubated for 24 hr in the absence or presence of DTT. No change in band pattern was observed, suggesting CDΔCwp84 SLPs contained no protease contamination, any contained proteases did not further alter the band pattern or SlpA does not auto-mature.

3.3.7 Ability of CDΔCwp84$_{347}$ to cause infection

To assess the capacity of CDΔCwp84 to cause CDI, the hamster model was used as it is the most widely recognized model currently available. This disease model parallels most of the recognized characteristics of the human disease and hamsters can readily be made susceptible to CDI by treatment with a broad spectrum antibiotic such as clindamycin. After administration of clindamycin, test groups were challenged 48 hr later by the orogastric administration of 2 x 10^2 spores produced from either CDΔCwp84$_{347}$ or wildtype *C. difficile* 630ΔErm.

Three experiments were conducted in the following manner; two experiments with animals challenged with CDΔCwp84$_{347}$ or wildtype *C. difficile* 630ΔErm and one with animals experiment challenged with just CDΔCwp84$_{347}$. In total 50 observations were made, 20 animals were challenged with the WT strain and 30 with the mutant strain.

A first experiment of 20 hamsters (2 test groups of 10) was challenged with wildtype *C. difficile* 630ΔErm or CDΔCwp84$_{347}$. In both test groups 90% succumbed to severe disease within 6 days post challenge. However, the difference in median survival time of 2.5 days for *C. difficile* 630ΔErm and 1day for CDΔCwp84$_{347}$ (mean survival time 3.6 and 2.4 days respectively) suggested slightly increased virulence of the CDΔCwp84$_{347}$ mutant (Figure 3.3.9 A).

To confirm this, a second experiment was performed but with only CDΔCwp84$_{347}$ challenged animals (1 test group of 10). Only 60% of challenged animals succumbed to severe disease within 6 days post challenge (Figure 3.3.9 B). The Kaplan-Meier survival curves appear more protracted and did not show the rapid onset of disease as previously observed, also reflected by an increase in the median survival time of 5.5 days. However, of the 60% that did succumb to disease, the median survival time was to 3 days, more consistent with the first experiment.

A final experiment was therefore performed to determine if the differences between the two experiments were consistent.

20 hamsters (2 test groups of 10) were again challenged with wildtype *C. difficile* 630ΔErm or CDΔCwp84$_{347}$ (2 control groups of 4 animals were unchallenged). 100% of the CDΔCwp84$_{347}$ and 80% of wildtype *C. difficile* 630ΔErm challenged animals succumbed to severe disease within 6 days post challenge. The difference in median survival time of 2.5 days for CDΔCwp84$_{347}$ and 4 days for *C. difficile* 630ΔErm confounds the possible increase in virulence of the CDΔCwp84$_{347}$ strain (Figure 3.3.9 C).

Unfortunately, due to technical issues two control animals also succumbed to symptoms of CDI, with one death. While faeces from challenge test groups isolated the respective challenge strain, the *C. difficile* strain(s) isolated from infected control animal(s) had a wildtype *cwp84* gene. It is unknown whether this strain(s) originated from other challenged *C. difficile* 630ΔErm animals or whether another *C. difficile* strain was brought in to the experimental environment.

Overall, of 30 hamsters (3 test groups of 10) challenged with CDΔCwp84$_{347}$, 83.3% succumbed to severe disease within 6 days post challenge compared to 85% challenged with *C. difficile* 630ΔErm (20 hamsters, 2 test groups of 10). Both strains displayed typical symptoms of CDI (diarrhoea, weight loss, lethargy). Figure 3.3.9D shows combined survival plots for this data which were analysed with a log rank test to assess the relative hazard of succumbing to disease. The difference in median survival time (with 95% confidence intervals) of 2 days for CDΔCwp84$_{347}$ and 4 days for *C. difficile* 630ΔErm was found not to be statistically different (p-value of 0.218, Log-rank test).

Comparison of the median survival times between the two *C. difficile* 630ΔErm and CDΔCwp84$_{347}$ experiments reveals that on average, CDΔCwp84$_{347}$ challenged animals die 1.5 days before WT challenged animals. Since only mutant-challenged animals were in Experiment 2, this might confound the relationship between challenge strain type and survivorship. When animals in experiment 2 are excluded, the hazard of dying for animals challenged with the WT strain is significantly lower (p=0.008, Log-rank test) than for mutant-challenged animals.

Faecal samples were removed from euthanised animals by colectomy and *C. difficile* isolated and analysed. PCR of the *cwp84* gene demonstrated that all *C. difficile* isolates from the CDΔ*cwp84*$_{347}$ test groups contained the expected intron (4.3 kbp product) while all isolates from the wildtype group gave a 2.4 kbp product consistent with the wildtype gene (as described in Figure 3.3.1A). The data demonstrates that the CDΔCwp84$_{347}$ mutant is capable of causing CDI in the hamster model at a similar rate of onset as the wildtype strain.

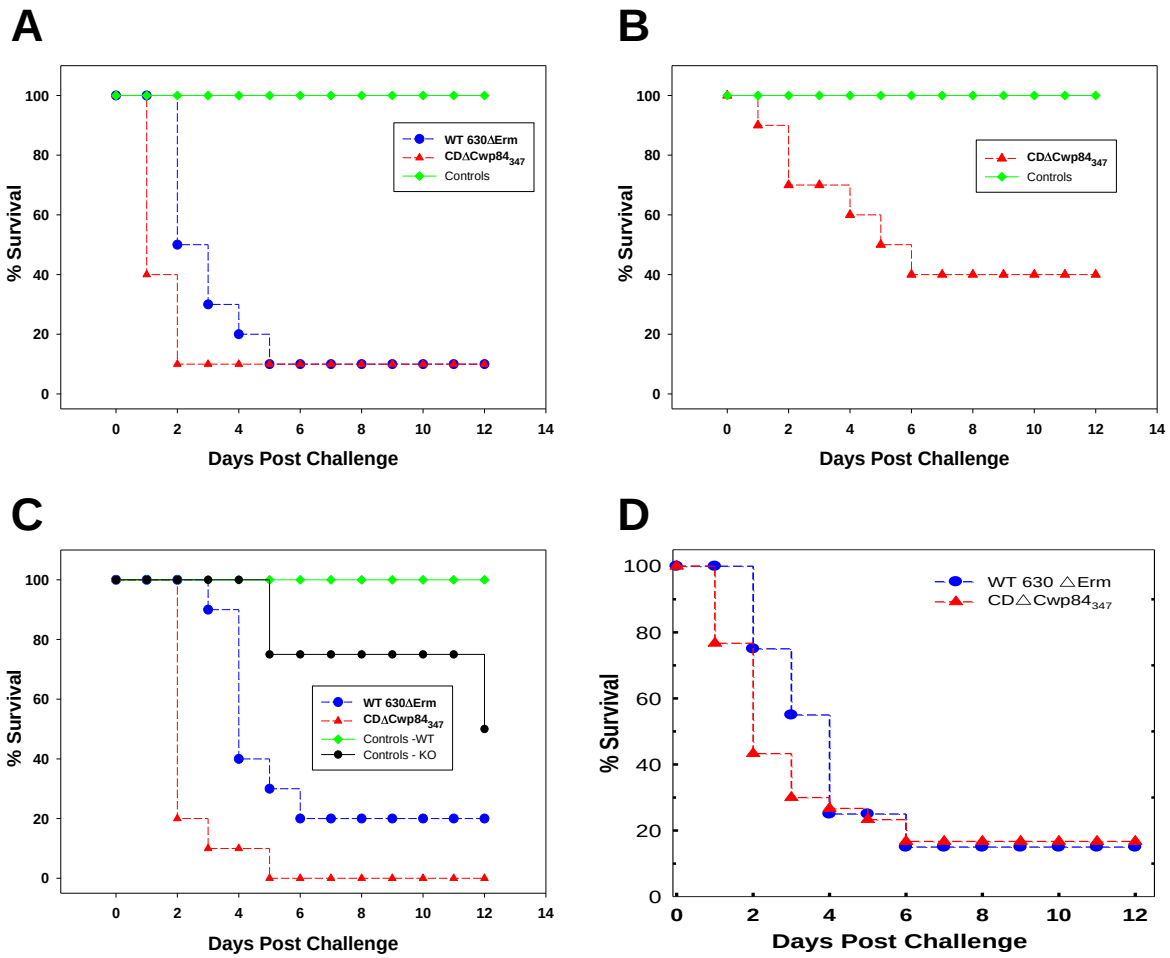


Figure 3.3.9 Virulence of *C. difficile* 630ΔErm and CDΔCwp84$_{347}$ in hamsters. (A) 2 test groups of 10 animals, challenged with CDΔCwp84$_{347}$ or *C. difficile* 630ΔErm. (B) 1 test group of 10 animals challenged with only CDΔCwp84$_{347}$. (C) 2 test groups of 10 animals, challenged with CDΔCwp84$_{347}$ or *C. difficile* 630ΔErm. (D) Combined Kaplan-Meier plots, for CDΔCwp84$_{347}$ the data represent 3 test groups of 10 animals and for *C. difficile* 630ΔErm, 2 test groups of 10 animals. Data shows time from challenge to severe disease/death in the hamster model for CDI.

3.3.8 Recombinant Cwp84

3.3.8.1 Cloning, expression, purification of rCwp84 and cleavage of SlpA.

Knockout of the *cwp84* gene results in the presentation of immature, i.e. uncleaved SlpA at the cell surface. The hypothesis stands that Cwp84 performs this maturation role and thus cleavage of SlpA may be demonstrable *in vitro,* i.e. with recombinant Cwp84. Recently, cleavage of a GST tagged SlpA fragment by rCwp84 has been demonstrated in *E. coli*, an action prevented by mutation of Cwp84's predicted active cysteine (C116A) (Dang *et al.* 2010).

To establish if rCwp84 could cleave SlpA *in vitro*, similar to the action of trypsin i.e. on extracted CDΔCwp84 SLPs, a number of recombinant Cwp84 variants were expressed, purified and their cleavage assessed. Table 3.3.4 summarises the Cwp84 variants produced, expression strains tried and the purified forms of full length (rCwp84) and cysteine protease domain ($rCwp84_{33\text{-}497}$).

Recombinant Cwp84 expression and purification has been described previously (Janoir *et al.* 2007) therefore the route for obtaining rCwp84 in this chapter mirrored this approach. Briefly, Janoir *et al.* (2007) expressed purified His_6-tagged full length Cwp84 (rCwp84) which upon incubation with DTT auto-matured to a 61 kDa (mCwp84) active species.

3.3.8.1.1 Initial problems

Initial efforts to express full length rCwp84 were hampered by loss of the DE3 prophage, particularly in liquid culture, and toxicity (from basal expression and upon induction). The use of only freshly transformed cells and the plating method of culture inoculation (Suter-Crazzolara & Unsicker 1995) allowed limited expression of pET constructs, as did use of *E. coli* K-12 derivative strains e.g. HMS174 (DE3) and Origami2 (DE3). Constructs which were either not under T7 polymerase control e.g. *tac* promoter, or which encoded only the cysteine protease domain of Cwp84 ($rCwp84_{33\text{-}497}$) were not so problematic. However, toxicity

putatively from expressing an active cysteine protease remained a problem, leading to lower cell yield and low recombinant protein yield.

Codon optimisation of the *cwp84* gene, for expression in *E. coli*, did not appear to alter construct stability/tolerance, as both natively cloned and synthetically synthesised (*E. coli* optimised) *cwp84* gene(s) behaved in a similar fashion. However, necessary mutations and modifications e.g. removal or stop codon(s) and changes in reading frame/restriction sites, were made to the optimised version of *cwp84*, as it was considered the most likely to give good expression.

3.3.8.2 Expression of full length rCwp84

3.3.8.2.1.1 *E. coli* expressed His-tagged rCwp84

Full length His_6 tagged variants of rCwp84 (cloned from *C. difficile* 630ΔErm or *E. coli* optimised) appeared to be expressed as an 84 kDa insoluble form at higher induction temperatures (30 and 37 °C) (as determined by western blot with anti-His_6 due to low expression levels). Changing media type or addition of media supplements (commercial or empirical) did not change expression levels.

When the induction temperature was reduced to 16 °C, 58 kDa (soluble) and 84 kDa (insoluble) bands reacted with anti-His_6 antibodies.

IMAC purification of soluble material expressed from the *E. coli* optimised construct yielded a doublet at 60, 58 kDa with a 10 kDa lower band, regardless of His_6 tag position. Material expressed from natively cloned *cwp84* (His_6.rCwp84) eluted impurely. However, western blotting suggested full length (84 kDa) rCwp84 was being expressed from T7.rCwp84.His_6 or natively cloned (His_6.rCwp84) constructs. The His_6 tags were removable from all N-terminally tagged versions with thrombin. However, incubation of all IMAC eluted soluble His_6 tag variants (pre or post thrombin treatment) with 5 mM DTT to assess self-maturation, did not affect band pattern on SDS-PAGE (Figure 3.3.10).

Table 3.3.4 Expression and soluble/insoluble analysis of recombinant Cwp84.

			Tag(s)			Expression 30°C (kDa)		Expression at 16°C (kDa)	
***cwp84* variant**	**source**	**vector**	**N-term**	**C-term**	**Expression strain**	**soluble**	**insoluble**	**soluble**	**insoluble**
Full length	630	pET28a	His		BL21 (DE3) Star	none	84	58	84
					T7 Express lysY/Iq	none	84	58	84
					IVT	84	84	n/a	n/a
	geneart	pET28a	His.T7		BL21 (DE3) Star	none	84	58	84
					Origami 2 (DE3)	58	84	58	84
					T7 Express lysY/Iq	none	84	58	84
					HMS174 (DE3)	58	84	ND	ND
					IVT	84	84	n/a	n/a
		pET24b	T7	His	BL21 (DE3) Star	none	84	58	84
		pET43.1a	NusA.His.S tag		BL21 (DE3) Star	62	multiple <130	as 30°C	as 30°C
		pET41a	GST.His.S tag		BL21 (DE3) Star	multiple <40	multiple <130	as 30°C	as 30°C
		pET41b	GST.His.S tag	His	BL21 (DE3) Star	multiple <40	multiple <130	as 30°C	as 30°C
		pGEX-6P-1	GST		BL21 (DE3) Star	84	84	84	84
		pGEX-6P-1	GST		BL21	84	84	84	84
$Cwp84_{33-497}$	630	pET28a	His		BL21 (DE3) Star	45	45	45	45
	630	pET28a	His		IVT	52 & 45	52 & 45	n/a	n/a
	geneart	pET SUMO	His.SUMO		BL21	none	65	65	65
	geneart	pTAC-MAT2		MAT	BL21 (DE3) Star	52^{Mut}	52^{Mut}	ND	ND

ND Not determined, n/a not applicable, ^{Mut} mutated version only (see Section 3.3.8.3.1.2

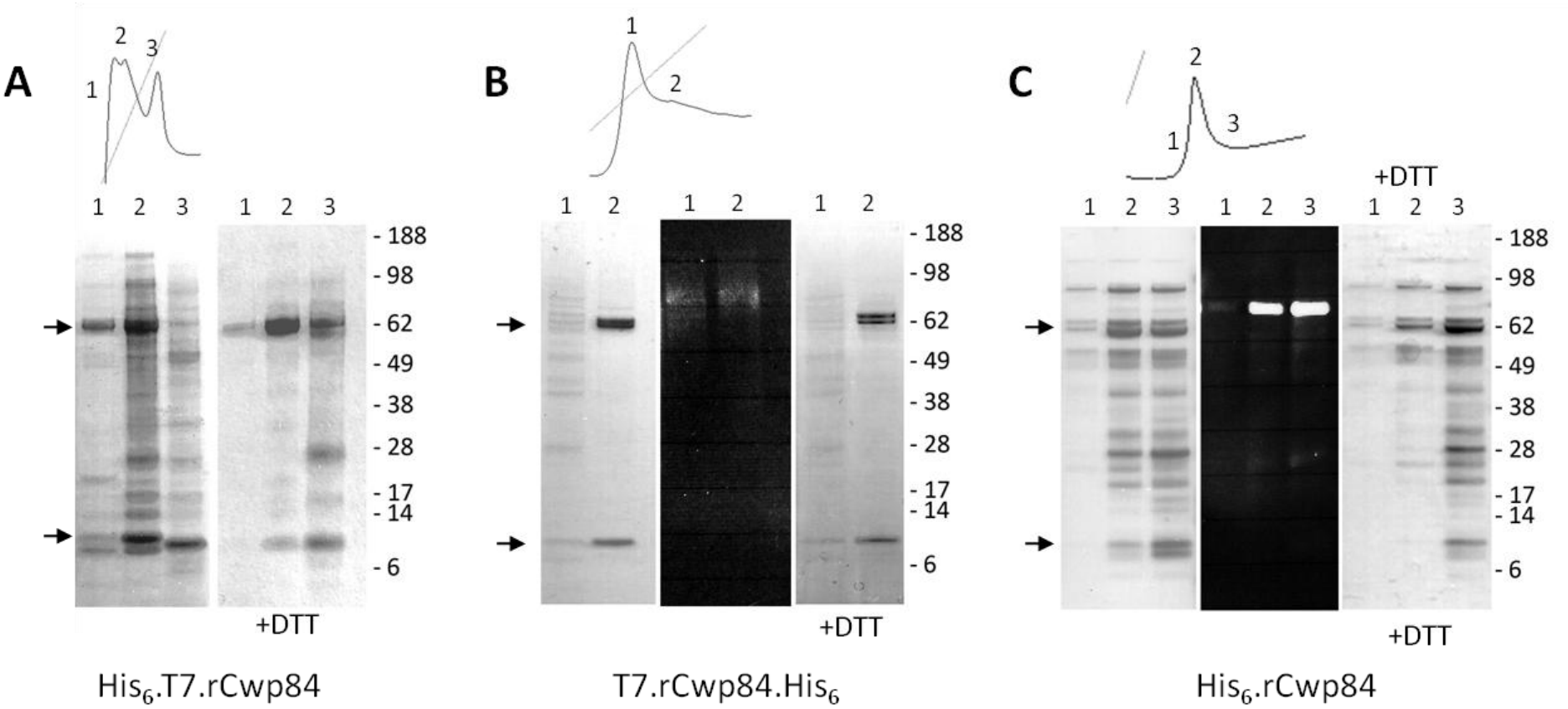


Figure 3.3.10 Analysis of soluble fractions of native IMAC purified His_6-tagged rCwp84 variants expressed at 16 °C (A) His_6.T7.rCwp84 elution chromatographs and associated eluted fractions, dialysed into PBS. (B) T7.rCwp84.His_6 chromatograph and associated eluted fractions & western blot with anti-T7 tag, fractions dialysed into 50 mM HEPES plus 0.15 M NaCl pH 7.4 (C) Natively cloned *cwp84* His_6.rCwp84 (630) chromatograph and associated eluted fractions, western blot with anti-His_6, fractions dialysed into PBS. After dialysis fractions were incubated with 5 mM DTT (in tris pH 8.0) for 4 hr.

Due to the truncation seen, inclusion body isolation and purification was assessed. Inclusion body isolation and denaturing IMAC of the His_6.T7.rCwp84 insoluble material eluted as a biphasic peak, which when analysed by SDS-PAGE gave three bands at 84, 76 and 66 kDa for both peaks (although peak overlap was pronounced) (Figure 3.3.11). Refolding of fractions from both of the His_6.T7.Cwp84 'peaks' by rapid dilution into a range of refolding buffers did not appear to alter the band pattern on SDS-PAGE. Furthermore, incubation with 5 mM DTT did not affect the band pattern (Figure 3.3.11). Remarkably, 84 kDa insoluble C-terminal His_6 -tagged rCwp84 did not appear to bind to an IMAC column in denaturing conditions.

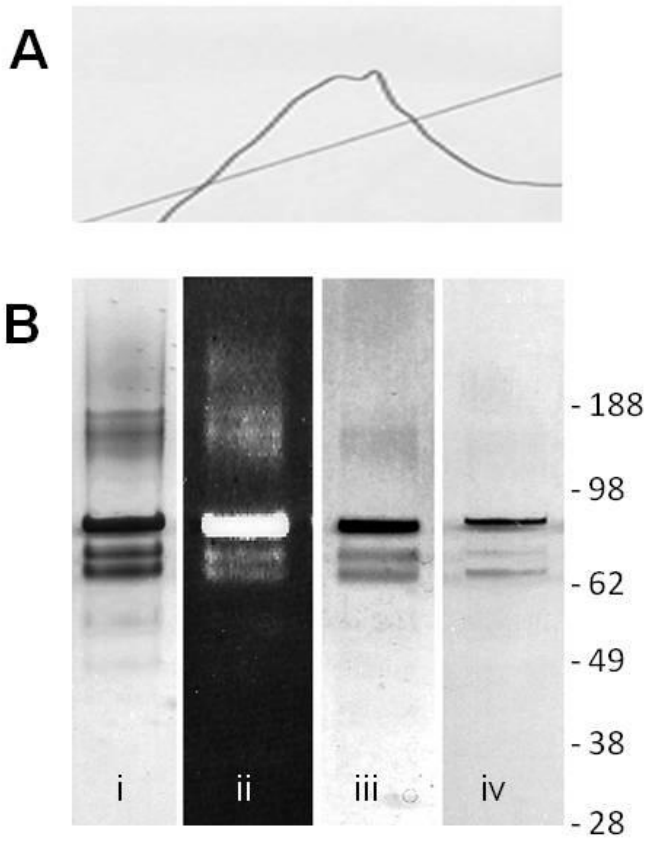


Figure 3.3.11 Inclusion body purification of rCwp84. (A) Denaturing IMAC chromatograph showing 'biphasic' eluted peak of His.T7.rCwp84. (B) (i) His.T7.rCwp84 denaturing IMAC eluted material. (ii) Immunoblot of IMAC eluted material with anti-His antibodies. (iii) His.T7.rCwp84 from (ii) after rapid dilution refolding. (iv) His.T7.rCwp84 from (iii) after 16 hr incubation with 2 mM DTT at 4 °C. Images representative of all refolding conditions and species eluted in IMAC.

3.3.8.2.1.2 Other fusion partners

In order to increase the expression of soluble rCwp84 and aid purification, a number of fusion tags were assessed, however none of the assessed fusion partners under either the T7 promoter (pET vectors with NusA or GST tags) or the *tac* promoter (pGEX-6P-1) produced satisfactory rCwp84. GST.rCwp84 (from a pGEX construct) appeared to be expressed as soluble and insoluble bands at 84 kDa, rather than the expected 119 kDa (84 kDa Cwp84 +35 kDa GST). The soluble 84 kDa did not show binding to glutathione resin and flowed through (suggesting GST tag occlusion) moreover incubation with PreScission protease for 4 hr did not cleave the GST tag from GST.Cwp84, again suggesting occlusion of the GST moiety. If however, GST.rCwp84 ran faster on SDS-PAGE (84kDa) than it should (93 kDa), then may resemble the His_6 tagged rCwp84 expression, i.e. a truncated 58 kDa rCwp84 species + 35 kDa GST tag.

3.3.8.2.1.3 *In vitro* transcription and translation (IVT) of His tagged rCwp84 (full length)

Expression of full length rCwp84 did not appear to be well tolerated in *E. coli*, potentially due to the expression an 'active' cysteine protease or misfolding and aggregation due to the CWBD interactions, particularly upon purification. To possibly overcome these problems, *in vitro* transcription and translation (IVT) of His_6 tagged rCwp84 constructs was assessed.

N-terminal His_6 tagged constructs appeared to express an anti-His_6 reacting 84 kDa protein in both total and soluble portions which purified using an IMAC micropurification procedure, although as a minor species (Figure 3.3.12). Incubation of the eluted material with DTT did not affect protein bands, pre- or post- His_6 tag removal with thrombin.

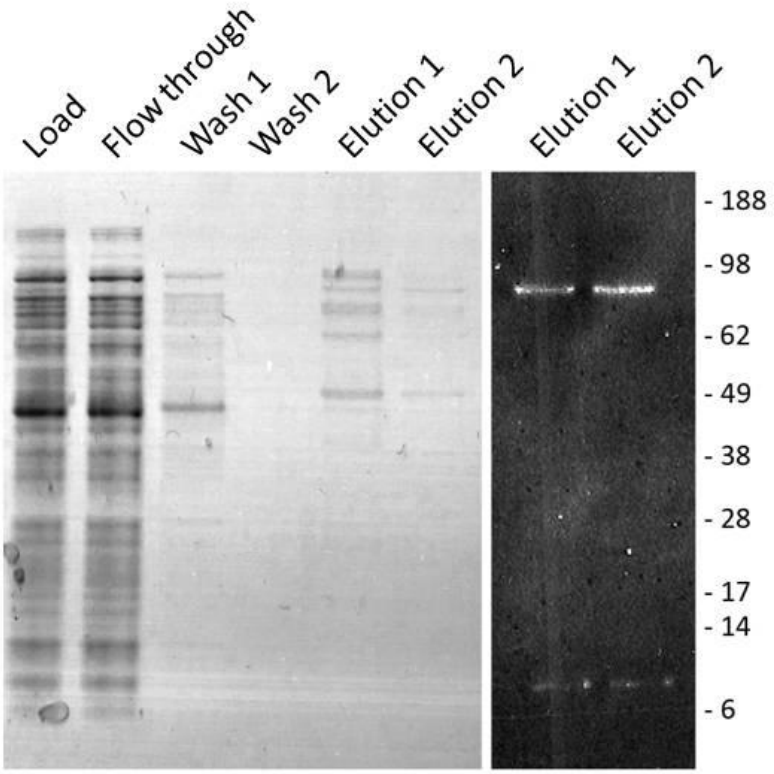


Figure 3.3.12 IVT purification of His.rCwp84 (630). Lanes 1-6, TALON IMAC micro-purification scheme from a 48 µl IVT reaction. Lane 7 & 8, Western blot on eluted material with anti-His_6.

In summary, only a 58 kDa soluble protein putatively rCwp84 could be purified from the full length gene expressed in *E. coli*. When expressed *in vitro,* a small amount of 84 kDa rCwp84 could be purified.

3.3.8.3 Expression of rCwp84$_{33\text{-}497}$ Cysteine Protease domain

The PFam 04122 cell wall binding domain (CWBD) is suggested to play a role in anchoring to, forming an array on (as assumed by the HMW SLP) or integrating with the cell surface. Recombinant expression of proteins containing these CWB repeats then are likely to retain this characteristic, thus during protein expression and purification, the protein of interest either aggregates with itself or other *E. coli* proteins possibly causing misfolding, and creates difficulties obtaining pure protein.

Therefore, defining and expressing the 'functional' part of the protein (without CWBD) is still likely to be valid and give important information about the business

end without interference from CWBD interactions. There is some precedent for this approach in the expression of SlpA where the LMW was soluble whereas the HMW (which consists of 3 x CWBDs) required refolding (Fagan *et al.* 2009).

Expression of just the cysteine protease domain (residues 33-497 (464 aa), predicted MW approx 52 kDa) was therefore assessed.

3.3.8.3.1.1 *E. coli* expression of rCwp84$_{33\text{-}497}$ using N-terminal His-tag

As there appeared to be little difference between *E. coli* optimised and natively (*C. difficile* 630ΔErm) cloned constructs, rCwp84$_{33\text{-}497}$ was expressed with only an N-terminal His$_6$ tag cloned from *C. difficile* 630ΔErm. The construct expressed as fragmented anti-His$_6$ reacting species at 52, 45, 42, 37, 28, 26 and 19 kDa. However, the anti-His$_6$ reacting species did not bind during IMAC purification.

Purification using IEX, as a first step, did not purify the main 45 kDa band away from contaminants, but appeared to increase the propensity to form 84 and 90 kDa species (Figure 3.3.13A). IEX purification at pH 6.5 (Q-sepharose) gave fractions which contained the same 45 kDa and 84 kDa species as pH 8.0, although slightly purer. This material when incubated with 5 mM DTT, resulted in the formation of a doublet at 42 and 45 kDa, other bands were unaffected (Figure 3.3.13B).

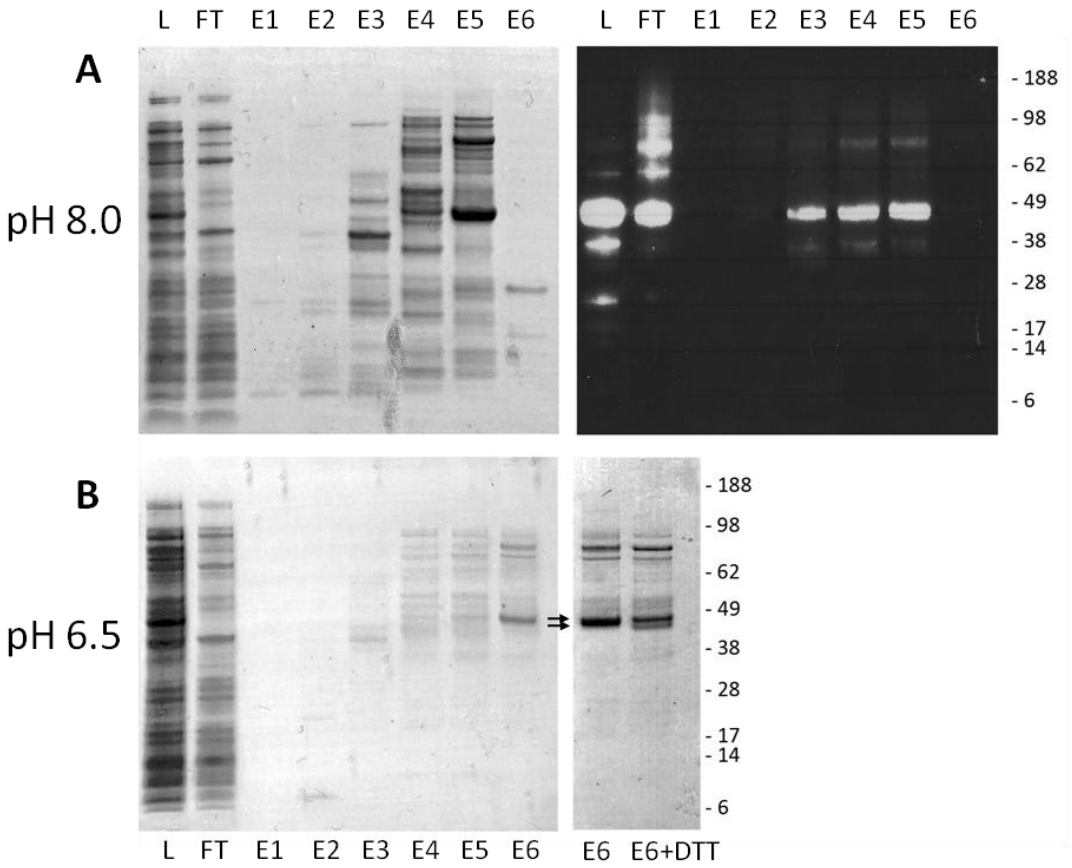


Figure 3.3.13 IEX purification of His.rCwp84$_{33\text{-}497}$. (A) Q sepharose (pH 8.0) purification fractions and western blot with anti-His. (B) Q sepharose (pH 6.5) purification fractions, including fraction E6 prior and post incubation with 5 mM DTT for 4 hr at 37 °C

3.3.8.3.1.2 *E. coli* expression of rCwp84$_{33\text{-}497}$ using C-terminal metal affinity tag

When rCwp84$_{33\text{-}497}$, with *E. coli* optimisation, was expressed with a C-terminal metal affinity tag (MAT) (HNHRHKH), as N-terminal tags may be removed during (auto)maturation, (rCwp84$_{33\text{-}497}$.MAT) a prominent 45 kDa protein (52 kDa predicted) was seen in the soluble (IMAC load) fraction but did not appear to bind in IMAC.

However, a rCwp84$_{33\text{-}497}$.MAT (S85G and T167I) mutant, obtained through errors during PCR amplification (identified by sequencing), expressed well and was easily purified as a full length 52 kDa protein (Figure 3.3.14). The 52 kDa S85G and T167I mutant rCwp84$_{33\text{-}497}$.MAT appeared to be unaffected by DTT i.e. could not

automature, but upon incubation with 5 µg/ml trypsin collapsed to a 46-49 kDa doublet i.e. being artificially matured by trypsin (similar to the effect of trypsin on rCwp84 (Janoir *et al.* 2007)). This data suggests that the mutation of S85G and T167I in rCwp84$_{33\text{-}497}$.MAT alters the protein structure such that it exposes the C-terminal MAT tag to allow binding to the IMAC resin and prevents automaturation, suggesting a requirement for correct structure for automaturation.

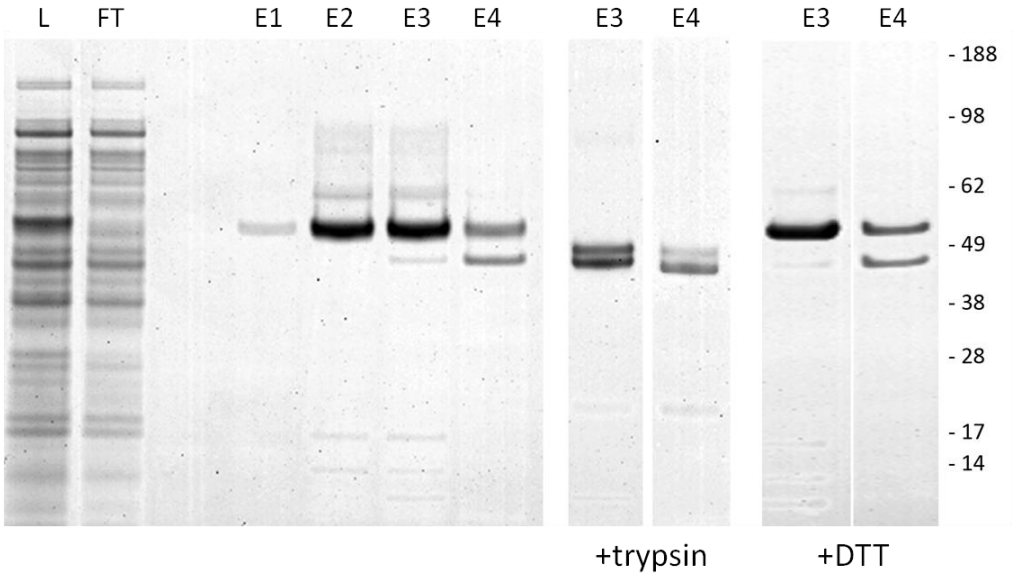


Figure 3.3.14 Purification of a C-terminal MAT tagged S85G & T167I rCwp84$_{33\text{-}497}$ mutant. IMAC purification and fractions E3 & E4 after incubation with 5 µg/ml trypsin or 2 mM DTT for 4 hr at room temperature.

The putative *in vivo* maturation of rCwp84$_{33\text{-}497}$ during expression in *E. coli* is highlighted by expression with an N-terminal His$_6$.SUMO tag. Western blotting IMAC fractions from His.SUMO.rCwp84$_{33\text{-}497}$ purification with anti-His$_6$ mainly detected two 23 and 25 kDa species, which would correspond to a His$_6$.SUMO tag (13 kDa) plus an N-terminal fragment of rCwp84$_{33\text{-}497}$, while full length His.SUMO.rCwp84$_{33\text{-}497}$ (65 kDa) and a 45 kDa species react much weaker (Figure 3.3.15 A).

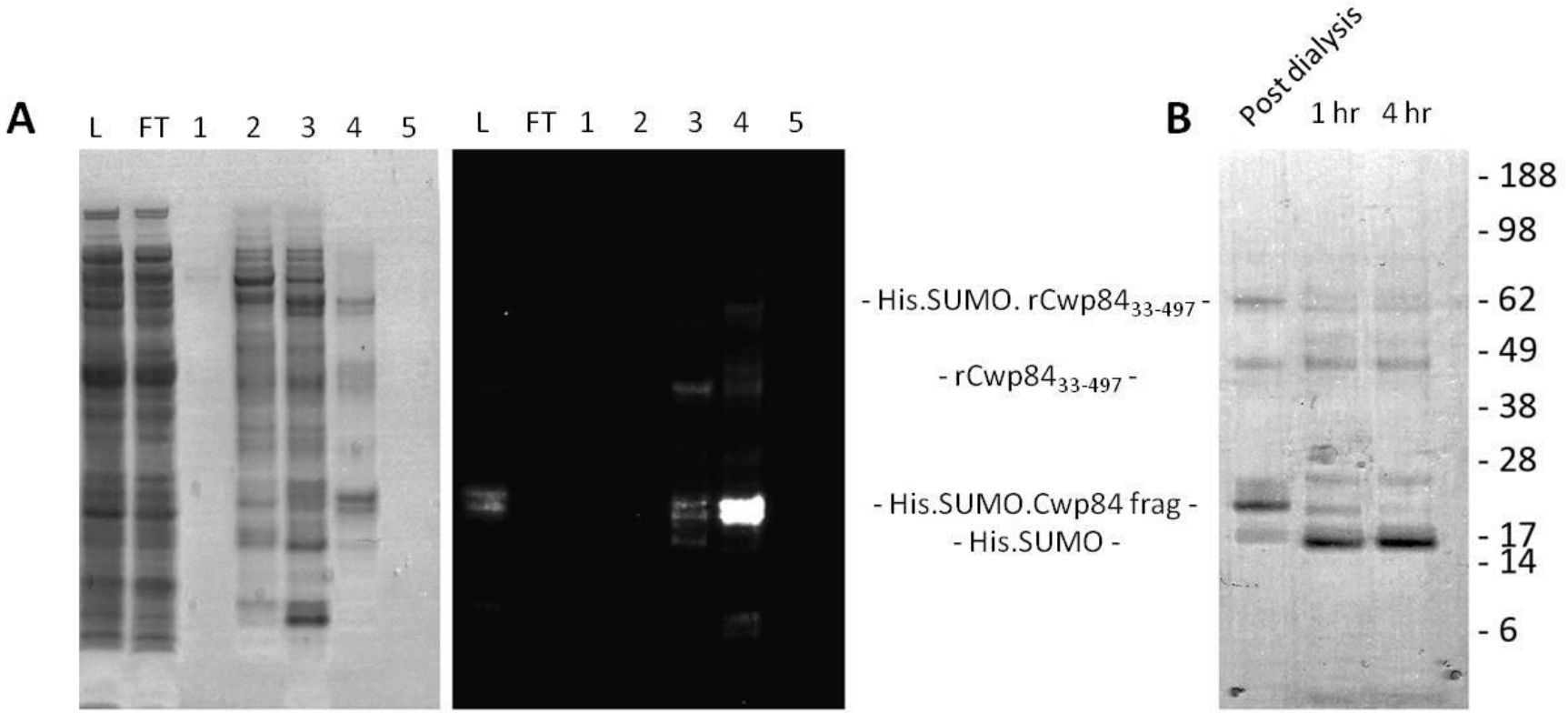


Figure 3.3.15 SUMO tagged rCwp84$_{33-497}$ purification and cleavage. (A) IMAC eluted fractions and western blot with anti-his. (B) Post dialysis of fraction 4 into SUMO protease cleavage buffer (50 mM tris-HCl, pH 8.0, 0.2% Igepal, 1 mM DTT) and incubation with 10 units SUMO protease at 30 °C samples taken at 1 hr and 4 hr.

The 23/25 kDa His_6.SUMO.rCwp84$_{33-497}$ fragment band is lost upon incubation with SUMO protease further confounding its identity. As expected, full length His_6.SUMO.rCwp84$_{33-497}$ (65 kDa) is cleaved leaving 52 kDa (rCwp84$_{33-497}$) and free 13 kDa His_6.SUMO (running at approximately 17 kDa on SDS-PAGE) (Figure 3.3.15 B). The apparent collapse of the IMAC eluted bands between 47-42 kDa (Figure 3.3.15 A Lane 4) into a band at 49 kDa after dialysis into the SUMO protease cleavage buffer (50 mM tris-HCl, pH 8.0, 0.2% igepal, 1 mM DTT) may also suggest automaturation under reducing conditions.

Taken together this data suggests 'maturation' of rCwp84 can occur in the cytoplasm of *E. coli* either through *E. coli* proteases or self-maturation. This maturation does not appear to be complete as multiple rCwp84$_{33-497}$ species are present.

3.3.8.3.2 In vitro transcription and translation of rCwp84$_{33-497}$

Anti-His_6 western blots of IVT expressed His_6.rCwp84$_{33-497}$ detect a 52 kDa (un-truncated) and a 42 kDa (presumably C-terminally truncated) species (Figure 3.3.16). The presence of full length His_6.rCwp84$_{33-497}$ in the IVT micro-purification elution fraction(s) suggests that, as found with IVT of full length rCwp84 (where full length 84 kDa found), material expressed using IVT has a decreased, but not completely abolished, tendency to automature or to be cleaved by proteases present either the IVT *E. coli* extract or in *E. coli* during 'normal' expression.

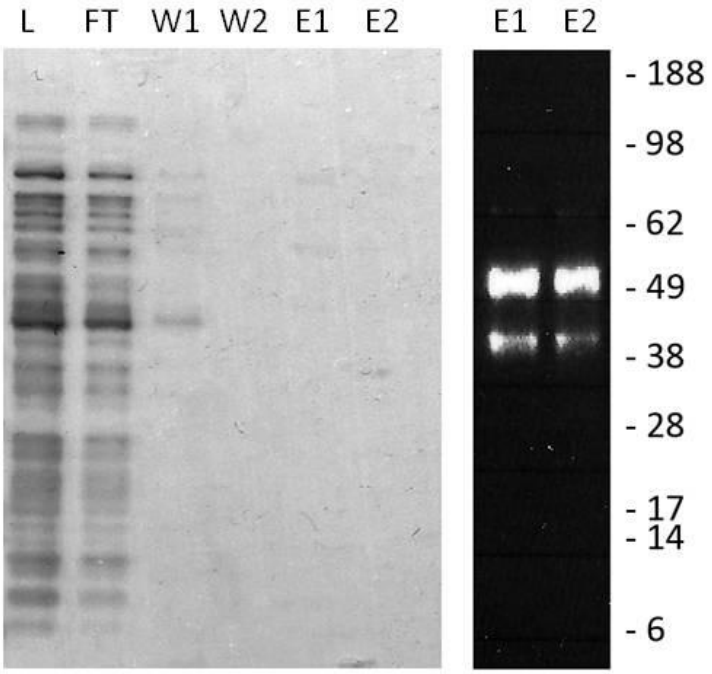


Figure 3.3.16 IVT purification of His.rCwp84$_{33-497}$. Lanes 1-6, TALON IMAC micro-purification scheme from a 48 µl IVT reaction. Lane 7 & 8, Western blot on eluted material with Anti-His$_6$. L- Load, FT- Flow Through, W1 & 2- Wash, E1 &2 - Elution.

3.3.8.4 Cleavage of SlpA with rCwp84

As aforementioned, the hypothesis stands that Cwp84 plays a role in S-layer (SlpA) maturation and thus the cleavage of SlpA may be demonstrable *in vitro*. Recombinant Cwp84 was therefore assessed for its ability to cleave SlpA *in vitro* as to date only the action of trypsin appears to cleave SlpA into bands resembling the WT HMW and LMW SLPs.

Obtaining recombinant Cwp84 of a sufficient purity restricted the variants of rCwp84 available for assessment. Table 3.3.5 summaries the rCwp84 forms assessed for their ability to cleave SlpA in CDΔCwp84 SLPs.

Table 3.3.5 Cleavage of CDΔCwp84 SLPs by recombinant Cwp84

Version		MW	Pre-reduced	Reducing conditions	Cleavage occurs	Effect on SlpA/SLPs	Inhibited by E64	concentration of rCwp84 (approx)	Time period
His.T7.rCwp84 (Insoluble)	*E. coli*	84,76,66	+	-	y	58 & 49 kDa	n/d	0.2 µg (200 ng)	5 hr
His.T7.rCwp84	*E. coli*	58	+	-	y	55,53+32	n/d	4 µg	o/n
	E. coli		-	+	n	-	n/a	10 µg/ml (0.2 µg)	5 hr
	E. coli		-	+	y	55,53+32	+	100 µg/ml (2 µg)	5 hr
T7.rCwp84.His	*E. coli*	58	+	-	y	55,53+32	n/d	1.5 µg	o/n
His.rCwp84	*E. coli*	84^{Δ}	-	+/-	n	-	n/a	-	
His.rCwp84	IVT	84^{Δ}	-	-	y	53	+	E1	o/n^{++}
			-	+	n*	36	-	E1	o/n
His.rCwp84		84^{Δ}	+/-	-	n*	36	n/d	E2	o/n
(His.)rCwp84		84^{Δ}	n/d	n/d	n	-	n/d	E2	o/n
His.T7.rCwp84	IVT	84^{Δ}	-	-	n*	36	n/d	E1	o/n
$His.rCwp84_{33-497}$	IVT	$52,42^{\Delta}$	-	+/-	n*	36	-	E1	o/n

All other rCwp84 species were either not pure enough or did not stick to column and therefore flow-through fractions were not considered pure enough

* the 38 kDa band disappears and a new 36 kDa band appears, unknown as to whether it is specific cleavage of SlpA

$^{\Delta}$ blottable species with anti-his

$^{++}$ reaction could occur faster than o/n (16 hr)

E (1 or 2) – Represent first or second washes of resin with elution buffer during micro-scale purification of IVT produced rCwp84 types (see section 3.2.1.4).

Protein concentration of the E1 or E2 fractions was not determined due to small volumes.

None of the recombinant Cwp84 variants assessed appeared to cleave SlpA in CDΔCwp84$_{347}$ SLPs into the 39 kDa and 48 kDa WT-like S-layer bands, as seen for trypsin, *in vitro* (Figure 3.3.17).

Full length refolded rCwp84 appeared to cleave SlpA into 58 and 49 kDa species (Figure 3.3.17 (i)), a result similar to cleavage of SlpA by low concentrations of papain (Figure 3.3.8). Full length soluble (58 kDa) and full length (cloned from *C. difficile* 630) IVT expressed rCwp84 cleaved SlpA in CDΔCwp84 SLPs into a 55,53 kDa doublet and a 32 kDa 'remainder'(Figure 3.3.17 (ii-v)). The demonstrated cleavage was inhibited by 100 µM E64, suggesting it was cysteine protease dependant.

In summary, a variety of recombinant Cwp84 variants were produced however none have shown cleavage of SlpA into WT-like HMW and LMW S-layer bands, as seen with trypsin. These data suggest that factors, or lack thereof, have prevented rCwp84 from cleaving SlpA *in vitro* to form WT like SLPs.

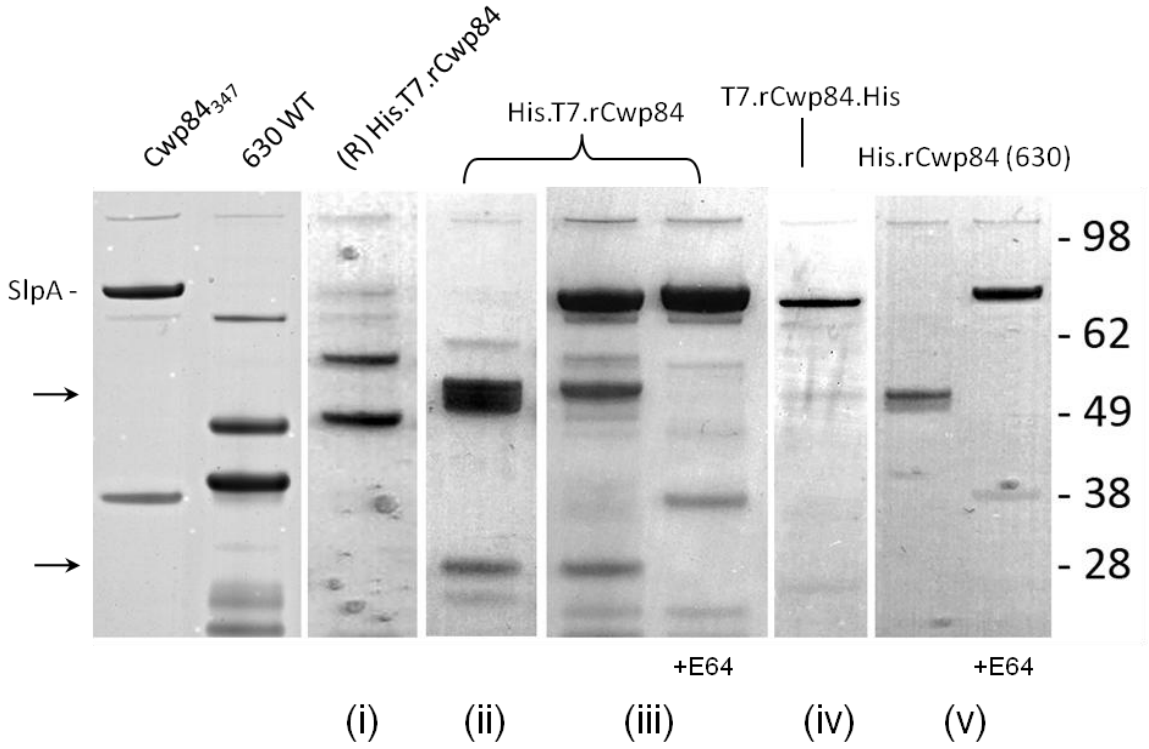


Figure 3.3.17 Cleavage of SlpA in CDΔcwp84 SLPs by rCwp84. Lanes 1 and 2 CDΔCwp84 & WT *C. difficile* 630ΔErm SLPs respectively. (i) Refolded (R) His.T7.rCwp84 cleavage of SlpA. 10 µl (~0.2 µg) refolded His.T7.rCwp84 (refolded by rapid dilution into 55 mM tris, 21 mM NaCl, 0.88 mM KCl pH 8.2 for 24 hr then incubated with 2 mM DTT for 24 hr) mixed with 10 µl (~3.45 µg) CDΔCwp84$_{347}$ SLP incubated at 37 °C for 5 hr. Cleavage in other refolding buffers was equal to or less than cleavage of SlpA in the above buffer system (ii) 16 hr incubation at 37 °C of 20 µl (6.9 µg) CDΔCwp84$_{347}$ SLPs with 20 µl (~4 µg) His.T7.rCwp84 in PBS pre-incubated with at 37 °C with 5 mM DTT for 4 hr. (iii) 2 µg His.T7.rCwp84 in PBS incubated with 20 µl (6.9 µg) CDΔCwp84$_{347}$ SLPs for 5 hr with addition of 5 mM DTT at 37 °C ± 100 µM E64. (iv) 16 hr incubation at 22 °C of 18 µl (6.2 µg) CDΔCwp84$_{347}$ SLP with 2 µl (1.5 µg) T7.rCwp84.His in 50 mM HEPES 0.15 M NaCl pH 7.4 pre-incubated for 4 hr at 37 °C with 2 mM DTT. (v) 16 hr incubation at 37 °C of 18 µl (6.2 µg) CDΔCwp84$_{347}$ SLPs with 2 µl IVT micropurified (elution 1) His.rCwp84 (630) in 50 mM sodium phosphate 0.3 M NaCl 150 mM imidazole ± 100 µM E64. CDΔCwp84$_{347}$ SLP extract in 50 mM HEPES 0.15 M NaCl pH 7.4.

3.3.9 Microscopy of Cwp84 mutant S-layer

Previous studies have demonstrated that the S-layer of *C. difficile* forms a regular array on the cell surface (Takumi *et al.* 1992; Cerquetti *et al.* 2000). By freeze etching whole cells and negative straining extracted SLPs, Cerquetti *et al.* (2000) suggested the presence of a dual layered S-layer; with a square ordered lattice on the outermost surface, with lattice spacing of ~10 nm, and a hexagonal lattice innermost. As KO of *cwp84* prevents S-layer maturation, a range of microscopy techniques were used to determine what effect the presentation of immature SlpA had on the cell surface, in particular if immature SlpA presented on the cell surface was able to form a regular array.

3.3.9.1 Cryo-SEM

Using low temperature-scanning electron microscopy (Cryo-SEM or LT-SEM), a method tending to replace freeze fracture replica techniques, no differences could be observed between the morphology of WT and CDΔCwp84 cells from solid phase bacterial culture (Figure 3.3.18). Examination of cells at low magnification revealed that CDΔCwp84$_{347}$ cells were surrounded by high levels of extracellular debris not present in 24 or 48 hr WT cells. Solid phase cultures were used to minimise physical stress and manipulation, e.g. centrifugation, as CDΔCwp84 is known to be sensitive to such treatments. Due to the slow growth of CDΔCwp84$_{347}$ only 48 hr cultures were used, although given the lack of stationary phase after 24 hr (Section 3.3.3.2), a similar outcome with cryo-SEM is likely. As no regularly ordered lattices were observed with either strain, other methods were pursued.

3.3.9.2 TEM

3.3.9.2.1 Whole cells

Examination of negatively stained partially lysed CDΔ*cwp84* cells by transmission electron microscopy (TEM) revealed an apparently regular parallel array on the cell surface, a portion of which had flapped over, yet retained this array (Figure 3.3.19 A). The array demonstrates an average lattice spacing of 7.5 nm (Figure 3.3.19 B).

TEM of whole *C. difficile* 630 ΔErm cells did not result in any discernable S-layer arrays only heavily stained cells (Figure 3.3.20).

3.3.9.2.2 Embedded sectioning

Embedded sectioning of *C. difficile* was first used by Kawata *et al.* (1984) to demonstrate the presence of the S-layer as two superimposed layers on the cell surface. Therefore to examine structures other than those directly on the bacterial surface, 24 hr cultures of both WT and CDΔcwp84$_{347}$ were embedded, sections taken then analysed by TEM (Figure 3.3.21).

Using this technique, it is demonstrated that *C. difficile* 630ΔErm possesses a cell envelope composed of six layers (Figure 3.3.21 A). The cell membrane innermost (layer 1) is followed by a light thin ~1.5 nm layer (layer 2), a ~2.5nm darker layer (layer 3) then a very lightly stained ~2 nm layer (layer 4). Following these is the densely stained (approx 15-30 nm thick) peptidoglycan (layer 5). On the outermost surface, and presumed to be the S-layer, are two layers: a lightly stained ~4 nm (HMW SLP) and a ~7 nm densely stained thick compact outer layer (LMW SLP) (layers 6(a) and (b) respectively). Comparision of CDΔcwp84$_{347}$ cells (Figure 3.3.21 B) with *C. difficile* 630ΔErm cells, reveals that layer 3 appears as a large ~6 nm densely stained layer, whereas this layer is only ~1 nm thick in the WT (Figure 3.3.21 A). The ill-defined appearance of the latter outermost layer (layer 6b), is consistent with the presentation of immature SlpA on the cell surface, which is possibly unable to form a properly locked S-layer array.

These results suggest that SlpA presented on the cell surface of CDΔcwp84$_{347}$ may lack the ability to form a regular packed lattice structure; moreover that CDΔcwp84$_{347}$ produces a significantly thicker, previously unidentified inner layer compared to the WT under similar growth conditions.

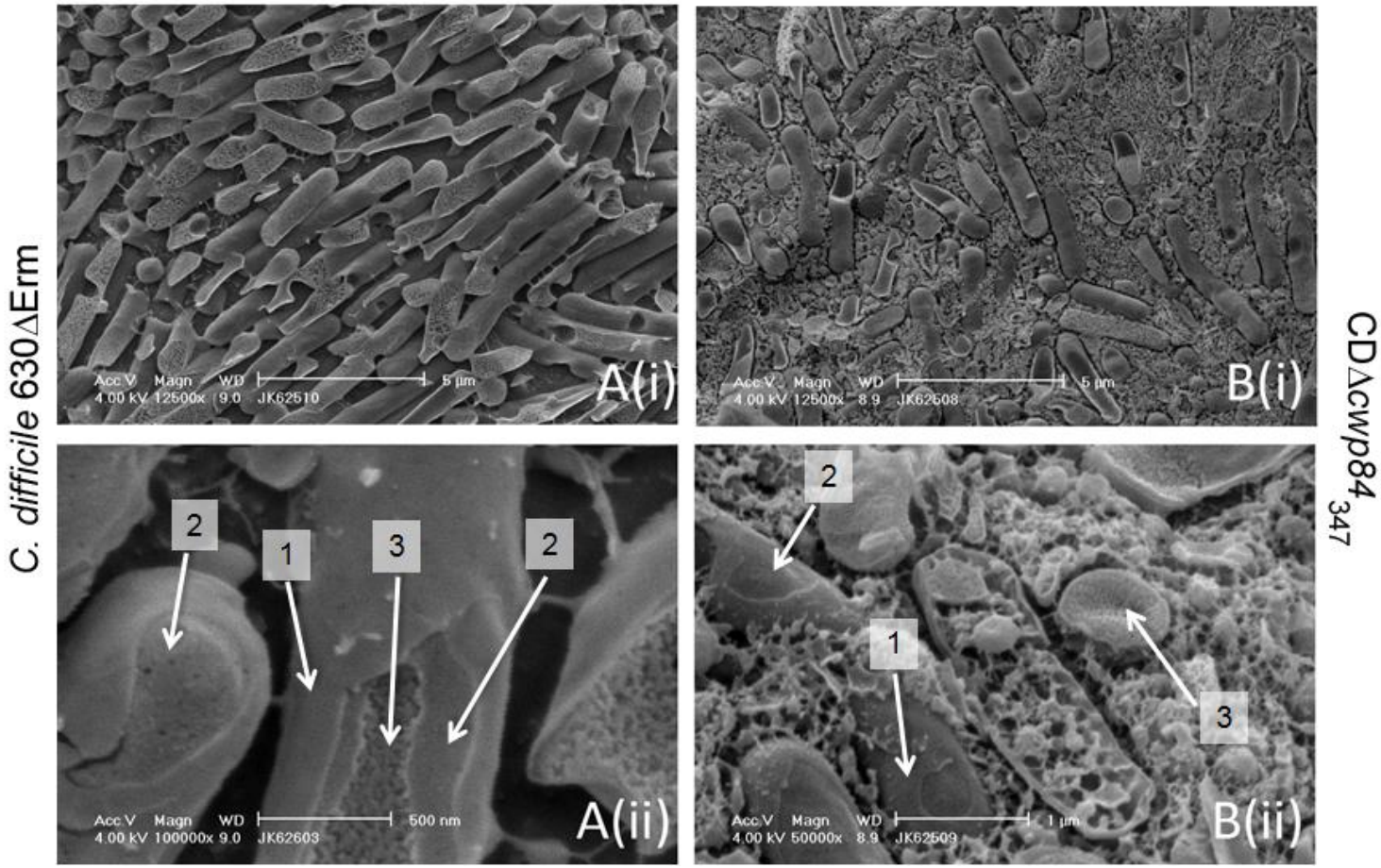


Figure 3.3.18 Cryo-SEM of *C. difficile* solid phase cells. (A) 24 hr *C. difficile* 630ΔErm cells at (i) low & (ii) high magnification (B) 48 hr CDΔCwp84$_{347}$ at (i) low & (ii) high magnification. Solid phase 48 hr CDΔCwp84$_{347}$ was used due to slow growth of the mutant compared to 24 hr *C. difficile* 630ΔErm.

KEY: 1 - Smooth outer layer, 2 - Perforated layer, 3 – Cell contents/cytoplasm.

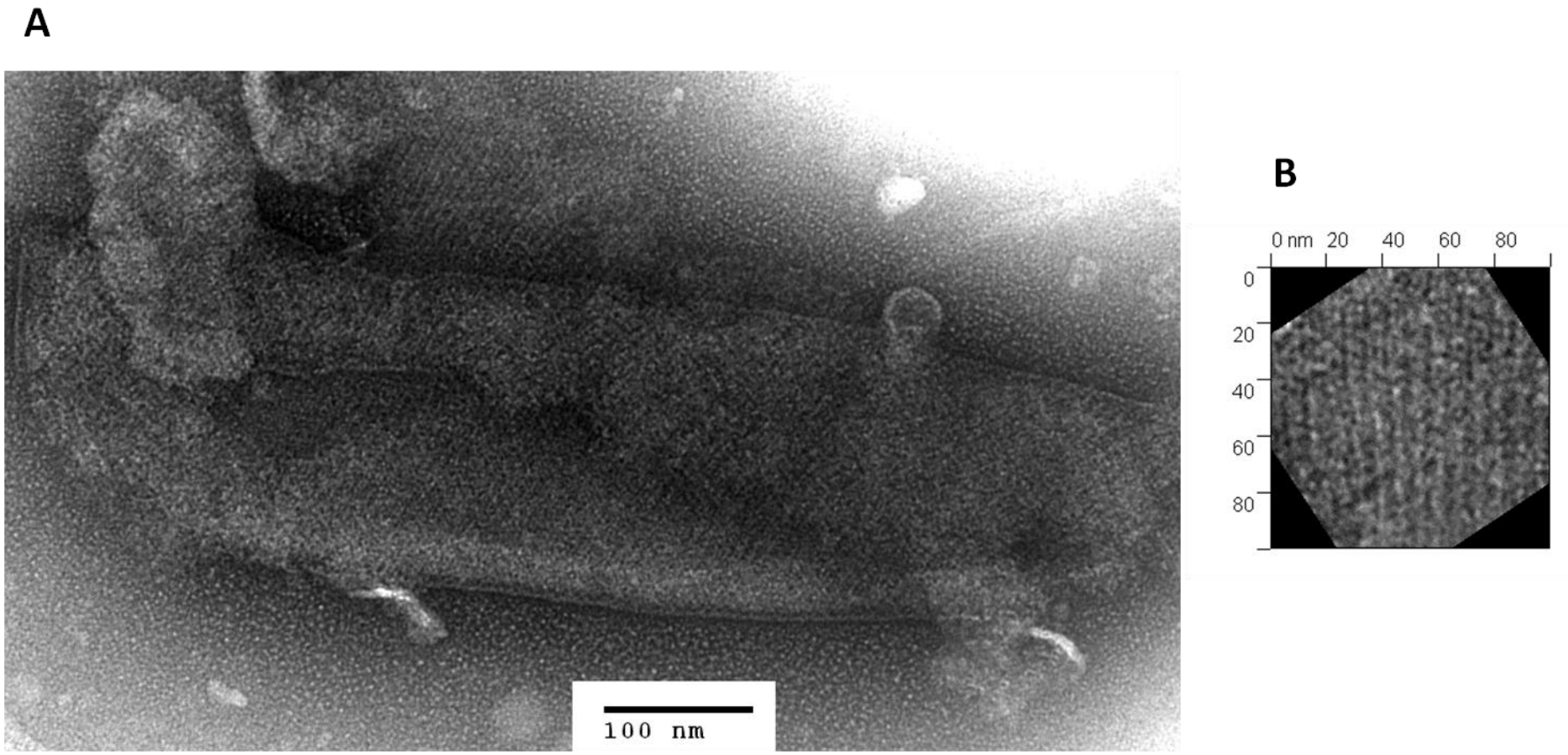


Figure 3.3.19 Demonstration of a regular array on CDΔcwp84 by TEM. (A) Negatively stained lightly sonicated CDΔ*cwp84* cell from 24 hr culture, revealing a regular parallel array (B) zoom in (A).

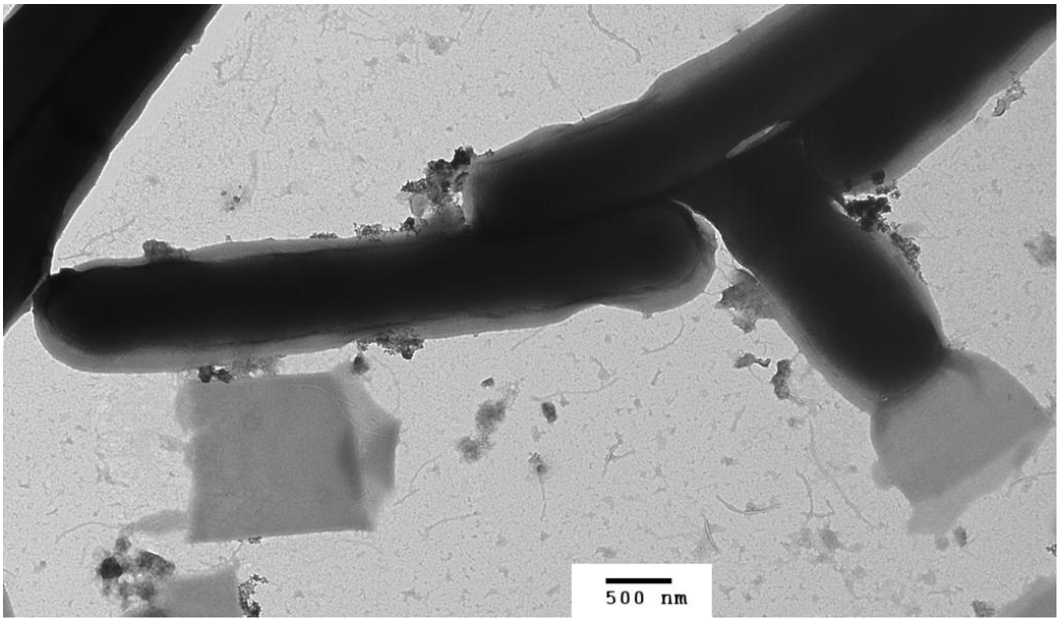


Figure 3.3.20 TEM examination of negatively stained, lightly sonicated, *C. difficile* 630ΔErm cells from 24 hr culture.

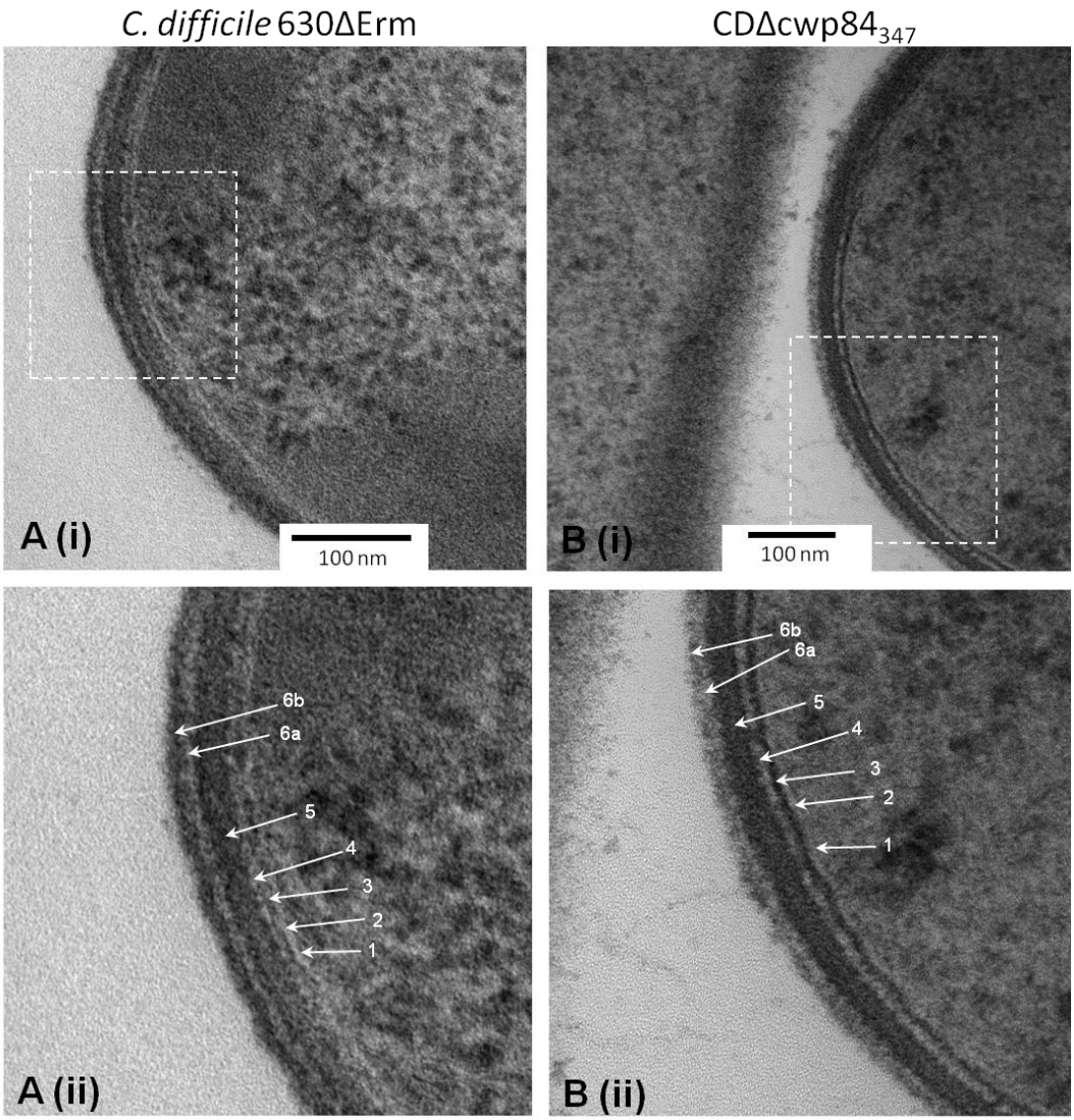


Figure 3.3.21 Embedded sections through *C. difficile* 630ΔErm and CDΔcwp84$_{347.}$ (A) wildtype *C. difficile* 630ΔErm and (B) CDΔcwp84$_{347}$. Image (i) acquired image, image (ii) electronically zoomed image of the area highlighted in image (i). Images are representative multiple grid examinations. Layer 1 - cell membrane, Layers 2-4- unknown, Layer 5 - peptidoglycan, Layer 6- S-layer (a) HMW SLP (b) LMW SLP.

3.3.9.3 AFM

Atomic force microscopy (AFM) is a powerful tool for visualising physical and chemical nanomaterials in their native environment. AFM is capable of imaging protein surfaces to nanometre level and S-layer arrays fall within this magnitude. AFM has been successfully used to map the surface of many bacterial S-layers including those which have been removed from the cell surface (Scheuring *et al.* 2002; Gyorvary *et al.* 2003; Dupres *et al.* 2009). AFM was therefore employed to help determine if immature SlpA in CDΔCwp84 SLPs affected S-layer array packing *in vitro* compared to the WT.

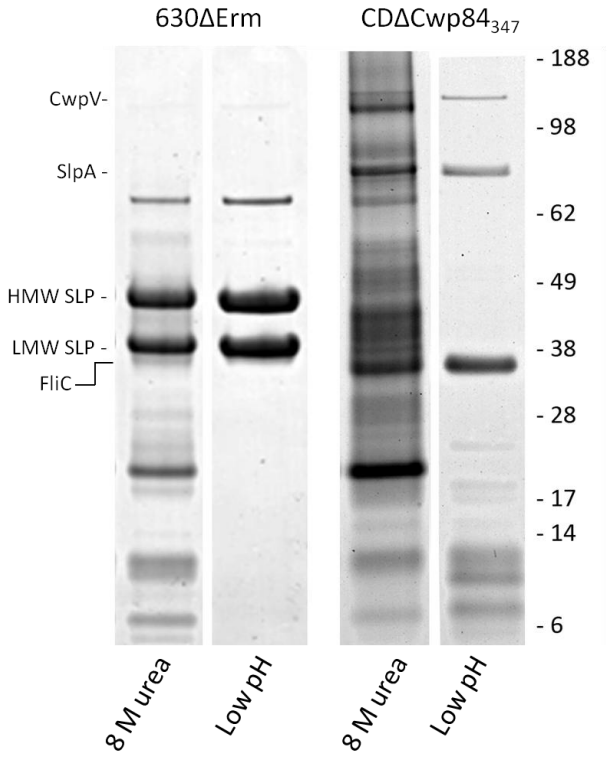


Figure 3.3.22 SDS-PAGE analysis of urea or low pH glycine extracted SLPs, from 24 hr cultures of *C. difficile* 630ΔErm or CDΔCwp84$_{347}$ used in atomic force microscopy.

Tapping mode AFM of WT 8 M urea SLPs (SDS-PAGE Figure 3.3.22), revealed that using a negatively charged (hydrophilic) surface (ozone modified graphite or particularly mica) the SLP extract formed a non-structured polycrystalline surface with raised plaque regions at low resolution (Figure 3.3.23 A). Upon closer

inspection of the plaques formed on mica, it was apparent that they exhibited a regular array (Figure 3.3.23 B). The array was interrupted by pores approximately 30-40 nm across. Examination of a 250 nm plaque region demonstrated that the array was made up of hexagonally packed units (Figure 3.3.23 C). The hexagonal arrays had a central core with a centre spacing of 100 nm, with approximately 100 nm spacing between each row of hexagonal units (Figure 3.3.25 A).

AFM imaging of the Cwp84 KO 8 M urea SLPs (SDS-PAGE Figure 3.3.22) required a similar substrate i.e. positivity charged (hydrophilic) surface but, also required the presence of calcium in the imaging buffer and could only be recorded using contact mode. Topographs of Cwp84 KO 8 M urea SLPs displayed a similar pattern of plaque formation, although plaques were not well formed (Figure 3.3.24 A). Closer inspection of a number of regions/plaques did not reveal any regular arrays (Figure 3.3.24 B). However, one 500 nm region clearly demonstrated a squarely arranged array (Figure 3.3.24 C). The tetragonal array in this region was characterised by a centre spacing of approximately 50 nm in both x and y dimensions, with a secondary array running alongside the main vertical planes, with a distance of 11-15 nm between it and the main array (Figure 3.3.25 B).

The AFM did not make contact with any surface when low pH SLP extracts were used, presumably because an array was not formed. This may have been due to impurities in the low pH SLP extract (despite a cleaner appearance in SDS-PAGE (Figure 3.3.22), a lack of component(s) present in 8 M urea extracts required for array formation or that other physical buffer/substrate conditions prevented array formation.

This data suggests that proteins in CDΔCwp84 8 M urea SLPs appear different *in vitro*, when examined using AFM, than the WT *C. difficile* 630ΔErm 8 M urea SLPs.

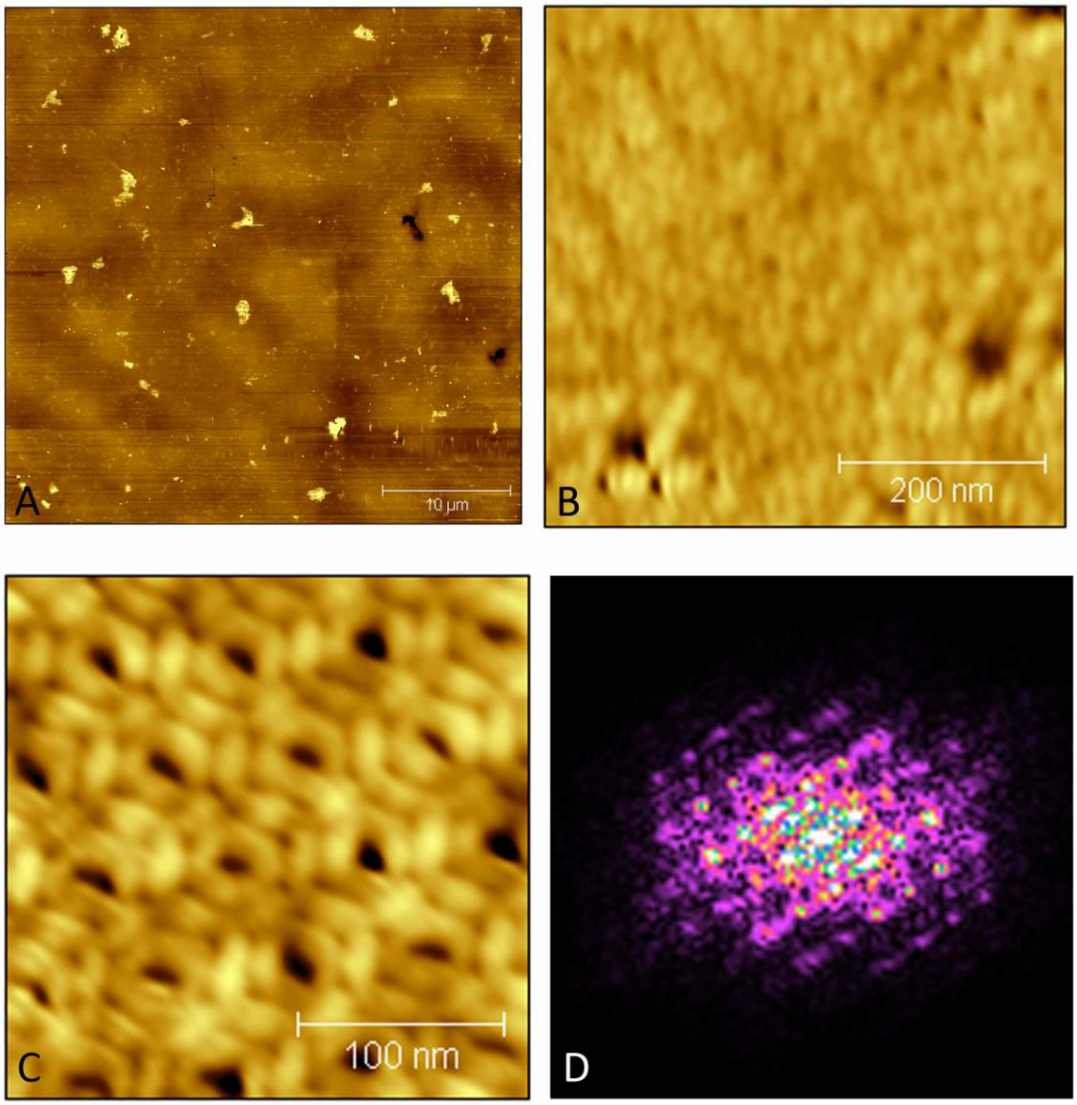


Figure 3.3.23 Tapping mode AFM topographs of *C. difficile* 630ΔErm 8 M urea extracted SLPs. (A) 40 µm scan region (Z-axis: 221 - 243 nm). (B) 500 nm scan of plaques seen in (A) (Z-axis: 0 - 23.9 nm). (C) 250 nm scan of (B) (Z-axis: 10 - 21.2 nm). (D) Fast Fourier transform (FFT) of (B). Imaging buffer 10 mM potassium orthophosphate adjusted tris buffer pH 7.8

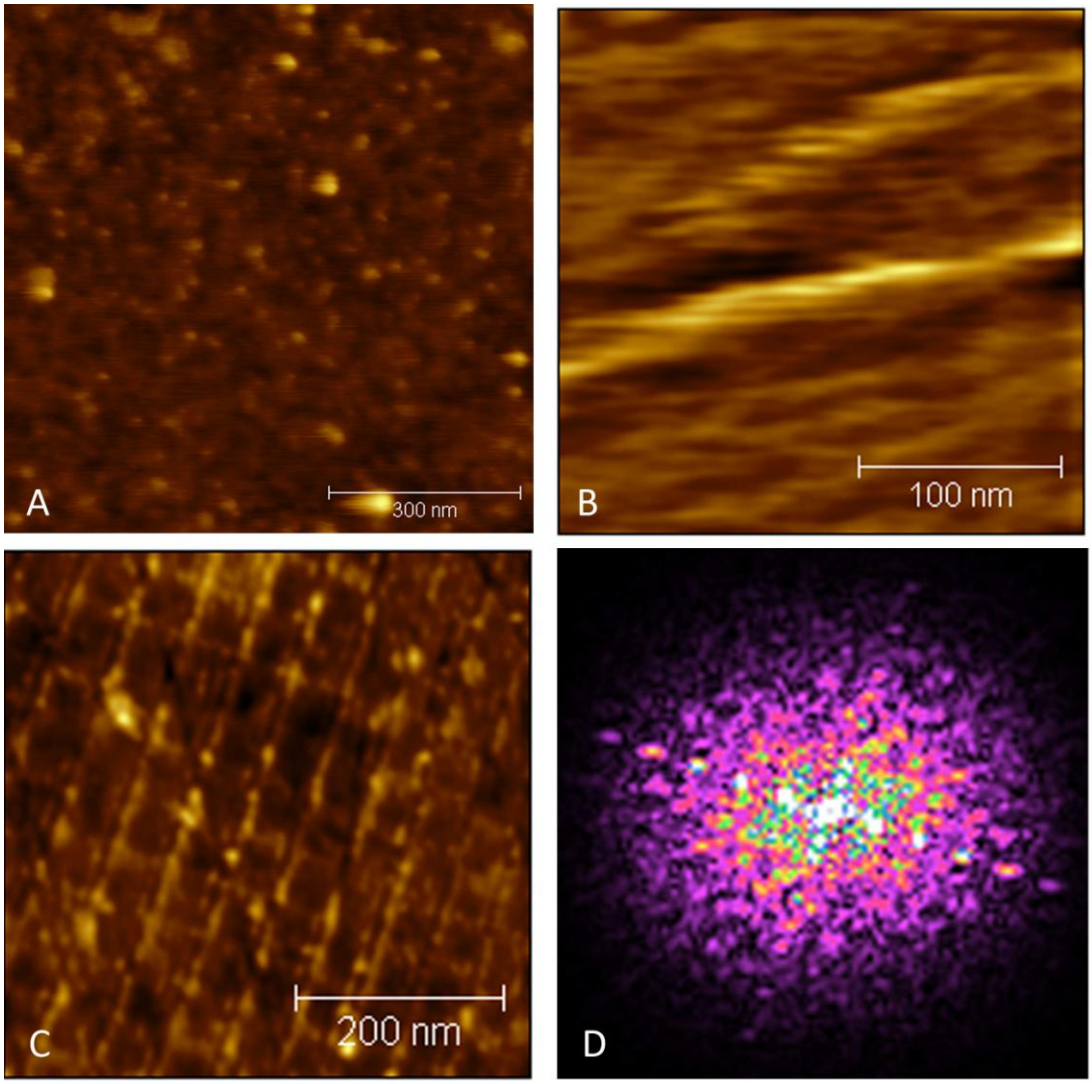
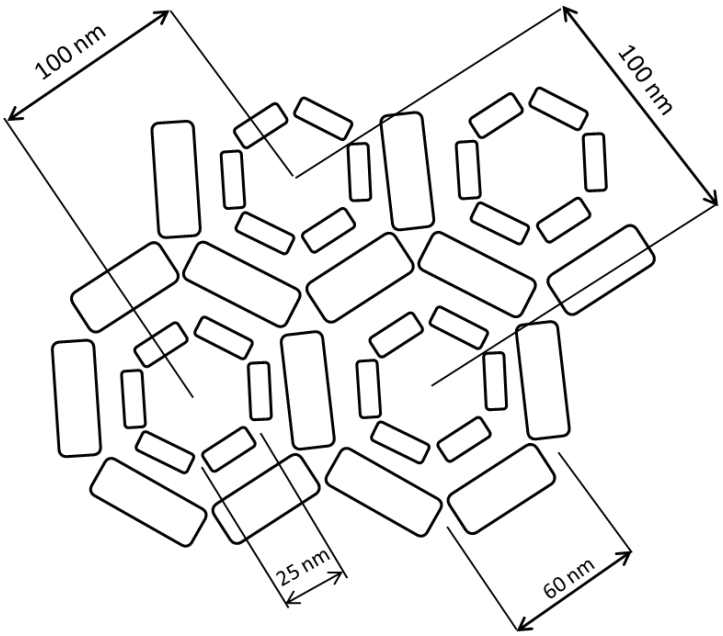


Figure 3.3.24 Contact mode AFM topographs of *C. difficile* ΔCwp84 8 M urea extracted SLPs. (A) 830 nm scan region (Z-axis: -3.5 – 9.6 nm). (B) 256 nm scan (Z-axis: -3.0 – 4.4 nm). (C) 500 nm scan (Z-axis: 0.0 – 20.0 nm). (D) Fast Fourier transform (FFT) of (C). Imaging buffer 10 mM potassium orthophosphate adjusted tris buffer pH 7.8 with additional 50 mM calcium chloride.

A



B

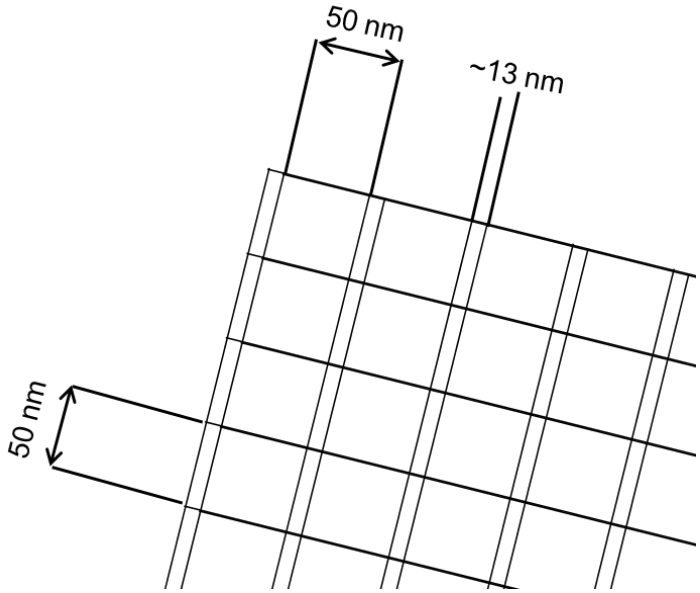


Figure 3.3.25 Diagrammatic interpretation of *C. difficile* 8 M urea AFM topographs. (A) Representation of hexagonal array seen in Figure 3.3.23(C). (B) Representation of tetragonal array seen in Figure 3.3.24(C).

3.3.10 Molecular modelling of the Cwp84 cysteine protease domain (Cwp84$_{33\text{-}497}$)

To understand more about structural aspects of Cwp84, tertiary structure prediction was undertaken with the N-terminal cysteine protease domain (residues 33-497 of Cwp84).

BLASTP (blast.ncbi.nlm.nih.gov) suggests approximately 44% of Cwp84$_{33\text{-}497}$ is similar, by sequence coverage, to other known cysteine proteases in the PDB. The highest hits are to human cathepsin S (2FYE and 2G6D, 30% identity each) and human cathepsin L (3OF8 and 3H89, 28% identity each).

Tertiary structure prediction by homology modelling is consistent with BLASTP results, as SWISSMODEL (Arnold *et al.* 2006) suggests residues 91 to 319 of Cwp84$_{33\text{-}497}$ have 21.4% identity with porcine cathepsin H, while CPHmodels (Nielsen *et al.* 2010) suggests 30.6% identity with human cathepsin S (2FYE). Superposition and active site comparison suggests the active site residues of Cwp84 are C116 and H262 (with Q110 and N261 assisting in orientation).

Fold homology structure prediction supports homology modelling, as the Phyre server (Kelley & Sternberg 2009) suggests fold identity with procathepsin B (3PBH) & procathepsin L (1CS8), HHPred (Soding *et al.* 2005) suggests identity with 3BPH and also procathepsin S (2C0Y)). Comparision of active site residues of the templates with the Cwp84 models also suggests C116 and H262 as the catalytic dyad (with Q110 and N261 assisting in orientation) (Figure 3.3.26).

HHPred and Phyre models are based on full length pro-enzymes i.e. still retaining the inhibitory pro-region. Analysis of the putative maximum pro-region (H33-N109) of Cwp84$_{33\text{-}497}$ models highlights how, consistent with other zymogens, the pro-region comes from above the active site cleft, through the 'oxyanion' hole and over one face of the enzyme into the bottom between the two domains (Cygler & Mort 1997) (Figure 3.3.26). Moreover, removal of a pro-region up to Q110, five residues from the active Cys116, would result in a ~12 kDa loss including the signal peptide

(~8 kDa without) in Cwp84. This highlights how a loss of 23 kDa, as suggested during maturation by Janoir *et al.* (2007), may be from other regions of the protein or is an artefact of recombinant production. Moreover, N287 (as proposed by Savariau-Lacomme *et al.* (2003)) does not appear to be close enough to interact with His262 in all models, suggesting either N262 performs this role, or a non-predicted fold brings N287 into play.

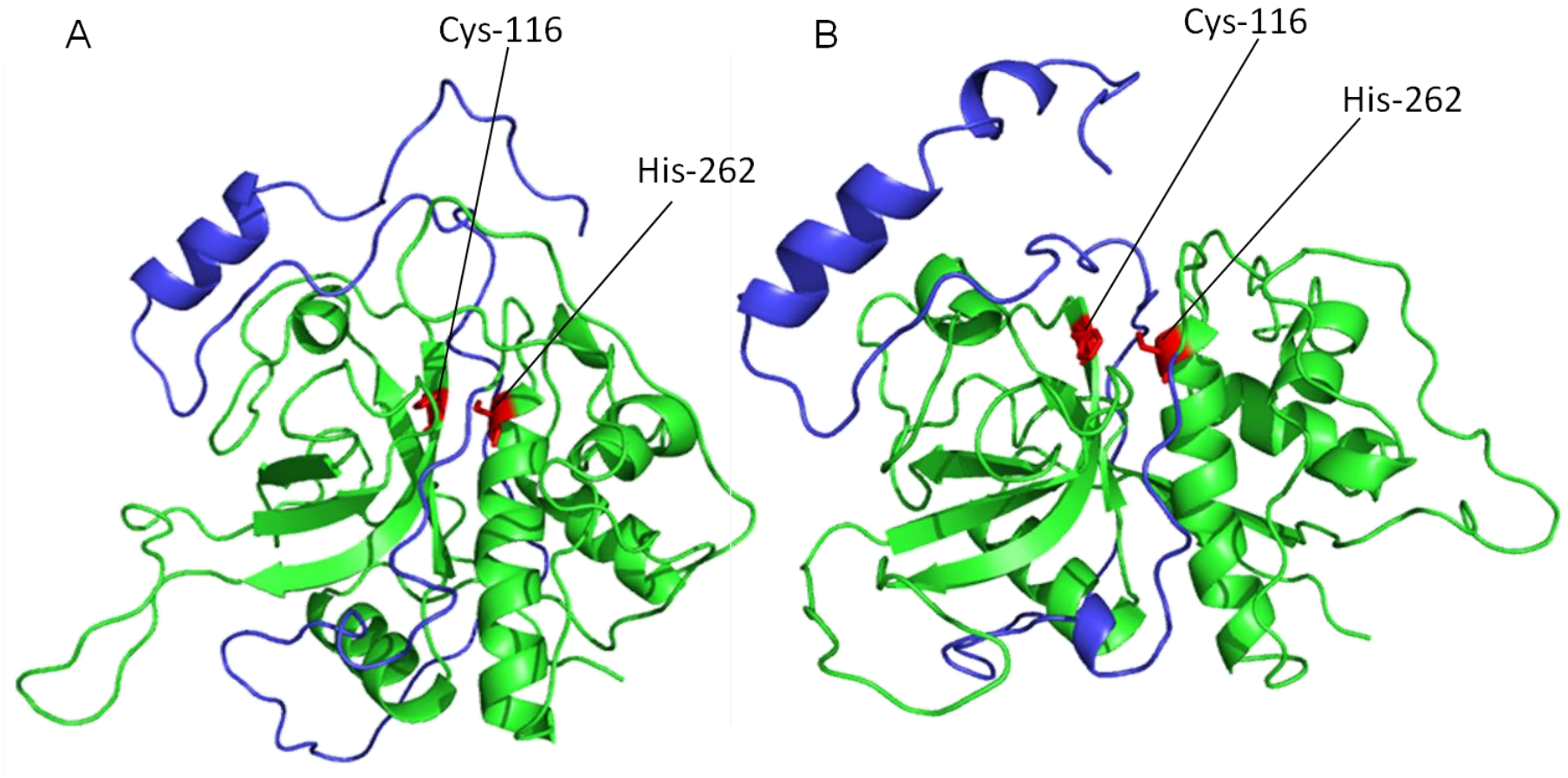


Figure 3.3.26 Tertiary structure prediction models of Cwp84$_{33\text{-}497}$ from Hidden Markov Model (HMM)-HMM comparison, by HHPred (Soding *et al.* 2005). Cwp84$_{33\text{-}497}$ modelled on (A) Procathepsin B (3PBH) (B) Procathepsin S (2C0Y). Active cysteine and histidine residues displayed in red, maximum proregion displayed in blue. Alignments modelled with MODELLER (Eswar *et al.* 2007) and prepared using PyMOL (www.pymol.org)

3.4 Discussion

It is only comparatively recently that it has become possible to generate stable, targeted gene knock-outs of *C. difficile* (Heap *et al.* 2007). Combined with proteomics and *in vivo* models for CDI, such techniques provide powerful tools with which to study disease pathogenesis. In this chapter, *cwp84* a gene encoding a putative surface protein Cwp84, has been knocked-out and it demonstrated that this cysteine protease plays a key role in the maturation of the surface protein layer.

3.4.1 Stringency of and genetic factors affecting KO outcome

While the KO of *cwp84* at multiple sites provides confidence that the phenotype observed was a result of the KO of *cwp84* and not promiscuous insertion. It is possible that the intron consistently inserted into not only *cwp84* but another gene which caused the resulting phenotype.

CD1751, a putatively surface located C1A family cysteine protease is theoretically a gene into which promiscuous insertion could occur. However, none of the insertion sites chosen for *cwp84* appear in *cd1751* and no screened clones had intron insertion into *cd1751*. The lack of a *cd1751* transcript and data by Dang *et al.* (2010) strongly suggests CD1751 does not play any significant role in pathogenesis. Inactivation of *cd1751*, for example using the ClosTron technology, may also confirm this.

Given that the CWBD is found in multiple copies in multiple surface proteins, intron insertion sites within CWBD(s) e.g. Cwp84 2054|2055a, could putatively insert into other CWBD(s). While CWBD identity is relatively low (~50%), if intron insertion sequences are sufficiently similar 'promiscuous' insertion is possible, insofar as any highly homologous sequence is a candidate for promiscuous insertion.

Group II introns integrate via a reverse transcription mechanism termed retrohoming, however a retrotransposition mechanism which occurs independently of the endonuclease function of the IEP, dependent on host recombinase functions,

has been found (Edgell *et al.* 2000) highlighting the possibility of intron mobility within CDΔCwp84. Complementation together with Southern blotting could be used to determine if, firstly returning the inactivated gene (*cwp84*) returns wild-type phenotypes and secondly confirmation that the intron has not entered at multiple sites in the target organisms genome. Studies using these analyses have been published (Heap *et al.* 2007; Emerson *et al.* 2009b) and it is feasible that until these analyses are completed, multiple copies of the ClosTron group II intron are present in CDΔCwp84 (or other mutants presented in this thesis). Surface inhibition of Cwp84 by Dang *et al.* (2010) also obtained an immature SlpA phenotype suggesting that the *cwp84* KO phenotype demonstrated in this chapter was due to lack of Cwp84 activity.

3.4.1.1 Genetic considerations of Cwp84 mediated cleavage of SlpA.

It is suggested that the ClosTron creates either a potentially active truncated product when an insertion site is at a distal location or a fore-shortened protein, still potentially retaining activity, when inserted at a proximal site (Heap *et al.* 2009a). Of the four sites chosen in *cwp84*, three (68|69a, 347|348a and 677|678a) were chosen such that any fore-shortened protein would be unlikely to retain any functional activity. Insertion site 2054|2055a, which would result in a fully translated N-terminal cysteine protease domain, still resulted in the same immature SlpA phenotype. This suggests that either the KO process (at any site, in any gene) alters the mRNA structure such that it is not translated, or that inactivation of the first CWBD of Cwp84 results in an inability to mature SlpA, suggesting that Cwp84 requires all the CWBDs for SlpA cleavage. The prominent species found on the cell surface is full length Cwp84 (85 kDa) (Janoir *et al.* 2007) and full length Cwp84 was found to cleave an SlpA fragment in *E. coli* (Dang *et al.* 2010).

3.4.2 Morphology and growth characteristics of cwp84 KO mutants

3.4.2.1 Morphological

All observations regarding changes in *C. difficile* colony morphology should be treated with caution due to the pleomorphic nature of *C. difficile* colonies,

specifically as the white and circular colony morphology of CDΔ*cwp84* has been observed in other *C. difficile* strains (Siani *et al.* (Reynolds *et al.* 2010; 2010)). Furthermore, *C. difficile cwpV* mutants exhibit a small and round colony morphology which can be only to the characteristic *C. difficile* colony morphology by complementation (Reynolds *et al.* 2010) Changes in colony morphology as a result of knockout of surface associated genes have been found in other species, for example S-layer (Rothfuss *et al.* 2006) and peptidoglycan hydrolase mutants (Camiade *et al.* 2010). Microscopic examination of CDΔ*cwp84* clearly shows WT cell morphology suggesting that a division septum is able to be formed (and cleaved). In CDΔcwp84, increased inter-cell adherence may inhibit movement across solid medium resulting in cell clumping rather than the typical spreading *C. difficile* colony morphology. Together with Cryo SEM of CDΔ*cwp84* data, this suggests the accumulation of extracellular material is likely to change both the refractive index of the colony (making colonies white rather than grey) and impede motility across the medium (causing round rather than spread colonies).

CDΔCwp84 grew more slowly and had a tendency to clump and form string-like precipitates, suggesting an increased degree of inter-cell adherence. In *Corynebacterium ammoniagenes*, a RamA (putative S-layer and cell wall biosynthesis regulator) knockout had increased cell surface hydrophobicity, leading to the formation of aggregated cell masses in liquid media (Lee *et al.* 2010b). Knockout of the majority of the *Thermus thermophilus* HB8 S-layer also lead to slow growth in liquid broth and a tendency to aggregate (Lasa *et al.* 1992).

A significant finding is the presence of immature SlpA in the growth medium during culture of CDΔCwp84. This suggests a weakened attachment via the cell wall-binding motifs, since the mature HMW SLP (or LMW SLP) was absent from the medium of wildtype *C. difficile* 630ΔErm. Immature SlpA presented on the surface of CDΔCwp84 is therefore unlikely to be able form a properly 'locked' lattice structure and, as a result, may be lost from the cell surface.

S-layer release into the culture supernatant has been described and may be connected with S-layer regulation through packing density (Breitwieser *et al.* 1992). Knockout of *cwp84* may therefore directly or indirectly de-regulate *slpA* (or SlpA), via a surface based feedback system. Sleytr & Glauert (1976) suggest that in *Clostridium thermohydrosulfuricum* and *C. thermosaccharolyticum,* during cell separation, a 'surplus' of S-layer subunits appears at the site of division thereby ensuring that the newly formed cell poles remain completely covered by the S-layer throughout the separation process. Excess (immature) S-layer presented on the cell surface of CDΔCwp84, beyond the normal 'surplus' required, could therefore be free to shed into the medium. Moreover, an S-layer transcriptional control (*slrA::kat*) mutant of *T. thermophilus* HB, like CDΔCwp84, grew slower and appeared to have trouble maintaining stationary phase (Bahl *et al.* 1997), highlighting how a loss of S-layer control results in slower growth.

3.4.2.2 Growth Rate

The knockout of *cwp84* leads to a reduction in cell integrity, primarily from having an altered S-layer. The observed slower growth of CDΔCwp84 is likely to be a result of an inability of the cell to maintain key physical characteristics normally upheld by an intact S-layer (Engelhardt 2007), i.e. cells are easily lysed. The *in vitro* action of trypsin on CDΔCwp84 SLPs, together with the observed increase in growth rate when trypsin supplemented culture media is used, suggests that trypsin is artificially maturing immature SlpA on the cell surface (or that which is shed into the culture medium) which is able to self-assemble and return some integrity to the cell. Therefore, trypsin causes the growth rate of CDΔCwp84 to appear to increase as lysis is being decreased.

It is also tempting to speculate that the increased inter-cell adherence in CDΔCwp84 may turn on quorum sensing systems that cause a range of effects including slower growth. Indeed there may indeed be a specific molecule or protein produced, or leached, from CDΔCwp84 which affects the bacterial cells in culture.

Other surface protein knockouts (See Chapter 5) also display slower growth rates (flagella cap (FliD) (p=0.053), Cwp66 (p=0.046) and CD2784 (p=0.027)), suggesting that other proteins are key in maintaining cell integrity or highlight a mechanism to decrease growth upon significant surface change.

3.4.2.3 Sporulation

Lawley *et al.* (2009b) found low levels of SlpA in the spore coat of *C. difficile* 630 spores. It follows that CDΔCwp84 spores are likely to be affected by the presence of immature SlpA polypeptide. However, CDΔCwp84 is able to a) sporulate, albeit at a reduced efficiency b) make viable, infectious spores. This suggests that for spore integrity it is not essential that SlpA be cleaved into HMW/LMW SLP or that, the domain(s) required for correct spore assembly are correctly folded. Decreased cell integrity is likely to lower sporulation efficiency, whereby spore formation creates stress on an already delicate cell resulting in cell lysis.

3.4.3 Analysis of surface layer proteins of cwp84 mutants

Extraction of the surface proteins of CDΔCwp84 primarily revealed the lack of SlpA maturation. However, a large number of other surface proteins are similar between the WT and Cwp84 KO indicating that these proteins are potentially unaffected by both *cwp84* KO and the resulting presentation of immature SlpA.

CwpV, a surface-associated protein which is also cleaved into two fragments (Emerson *et al.* 2009b) was found matured, as normal, in the SLPs of CDΔCwp84 suggesting a different/additional processing mechanism for this surface protein. A band of similar size is found in the culture supernatant of CDΔCwp84, suggesting CwpV is also lost to the culture supernatant, possibly as a result of an association with SlpA or that CwpV's regulatory control is altered resulting in excess CwpV and subsequent release into the medium.

3.4.4 Identification of Cwp84 at the cell surface

Proteomic analyses in this study and in a previous study by Wright *et al.* (2005) have not identified Cwp84 (aka Paralog 4 in Wright *et al.* (2005)) in glycine extracts of *C. difficile* 630(ΔErm in this study). However, Janoir *et al.* (2007) using antibodies

raised against matured rCwp84, and Dang *et al.* (2010) using pull down assay of ABP labelled Cwp84, have both found Cwp84 in glycine extracts.

The inability to identify Cwp84 in glycine extracts in this study could be due to the variability of batch cultures and the culture growth conditions. The Cwp84 may be highly unstable in the extraction conditions. The non-identification of Cwp84 in glycine extracts could, however, be due to limitations with a mass spectrometry approach to protein identification in complex mixtures (Baldwin 2004).

Antisera raised against peptides to predicted epitopes were unsuccessful at finding Cwp84 in glycine or lysozyme extracts. The primary reason for this is assumed to be the insolubility of the lyophilised peptides (conjugated to KLH). PBS ± 10% DMSO ± sonication did not appear to completely dissolve the conjugated peptides. Therefore, the amounts of soluble peptide immunised may have been far lower than expected, and the resulting concentrations of Cwp84 peptide specific antibodies may have been low. However, there is a possibility that usable antibodies were generated and the negative immunoblots were valid, as there may have not been any Cwp84 present in the glycine or lysozyme cell wall fraction(s). This may be due to the culture conditions, genuinely actual low (undetectable) amounts of Cwp84 or due to protein fragility/instability during the extraction procedure.

As recombinant Cwp84 (full length-truncated or N-terminal domain only) became available, this should have been used for antiserum generation, comparable to the method used by Janoir *et al.* (2007). Alternatively, a transcriptomic approach (RNA-seq) may have identified full length *cwp84* mRNA and its subsequent truncation in CDΔCwp84. Given Cwp84's key role in S-layer maturation there should be a ready supply of its message for RNA-seq, despite its putative short half-life to ensure Cwp84 is tightly controlled.

3.4.5 Effect of proteases on SlpA *in vitro* and effect on Cwp84 mutants

Incubation of CDΔCwp84 SLPs with trypsin cleaves SlpA leaving bands that are indistinguishable in size to the WT HMW and LMW S-layer subunits. Moreover, SLPs extracted from CDΔCwp84 grown in trypsin supplemented media are also cleaved to WT-like bands. This indicates that SlpA in CDΔCwp84 SLPs and immature SlpA presented on the surface of CDΔCwp84 are structurally similar and spatially in as much as the trypsin cleavage site (TKˇSANDT) is accessible.

Incubation of CDΔCwp84 SLPs with other selected proteases, including papain and cathepsin B from the peptidase C1A family (to which Cwp84 is proposed to belong), did not cleave SlpA in CDΔCwp84 SLPs to yield WT-like bands. It is unknown as to whether this data merely highlights the specificity of the assessed proteases cleavage residue(s) (or lack of for papain) or a 'unique' spatial orientation/structural conformation between trypsin and SlpA. However, full length rCwp84 was found to cleave an SlpA fragment (amino acids 1-362, fused to a C-term GST tag) rather than full length SlpA (when co-expressed in *E. coli*) (Dang *et al.* 2010), suggesting the structure of full length SlpA *in vitro* affects cleavage (even by Cwp84) into the respective SLP subunits.

In addition to the cleavage of SlpA in CDΔCwp84 SLPs, there are two other interesting points. First, the 120 kDa protein CwpV appears to be resistant to degradation by a number of the proteases tested in this chapter, suggesting stability to high (digestive) proteolysis potentially key *in vivo*. Secondly, *in vitro*, the HMW SLP appeared more sensitive to trypsin degradation. Early work on the S-layer of *C. difficile* highlighted how a 32 kDa protein, subsequently named as the LMW SLP, was resistant to trypsin but not pepsin (Kawata *et al.* 1984; Takumi *et al.* 1987). This may support the proposed orientation of the S-layer in which the HMW SLP is the lesser exposed of the two subunits (Fagan *et al.* 2009), leaving the more protease resistant proteins, e.g. LMW SLP subunit and CwpV, to confer a protective coat to the bacterium.

3.4.6 Ability of CDΔCwp84$_{347}$ to cause infection

CDΔCwp84$_{347}$ was assessed for its ability to cause CDI in the hamster disease model. The mutant strain was clearly competent at colonising the hamster gut and caused a similar spectrum of symptoms to wildtype *C. difficile* 630ΔErm with mean of fatality of 83.3% in the test groups. Bacterial isolates from the faecal samples taken from diseased animals confirmed CDΔCwp84$_{347}$ as the causative strain and further illustrated the stability of the mutation. These data demonstrate firstly, that Cwp84 does not directly play a critical role in CDI pathogenesis in the animal model and secondly, that bacteria expressing immature SlpA are still competent at causing disease. With respect to the latter, one possibility is that the SlpA precursor polypeptide retains any key biological action(s) of the HMW and LMW SLPs, and that the domains mediating this activity are correctly folded within the SlpA precursor.

Another possibility is that the SlpA precursor is 'matured' via cleavage by the gut proteases, as proven by cleavage with trypsin. Thus, it is possible that a proportion of CDΔCwp84$_{347}$ growing in the gut derives a mature S-layer via the action of gut proteases. Alignment of 14 SlpA sequences (Eidhin *et al.* 2006) at their putative cleavage sites suggests that the SlpA precursor could potentially be cleaved by gut enzymes trypsin or chymotrypsin (Figure 3.4.1). Examination of the gut lumen contents after challenge, which would include all bacterial (predominantly *C. difficile*) species and identification of an appropriate size band by western blotting with anti-LMW SLP may confirm SlpA cleavage by gut proteases. It remains true that *C. difficile* expressing the SlpA precursor polypeptide are still competent at causing disease in the hamster model. The ability of CDΔCwp84$_{347}$ to cause infection may be assisted by changes in toxin production, which may negate the deleterious effects of an immature S-layer.

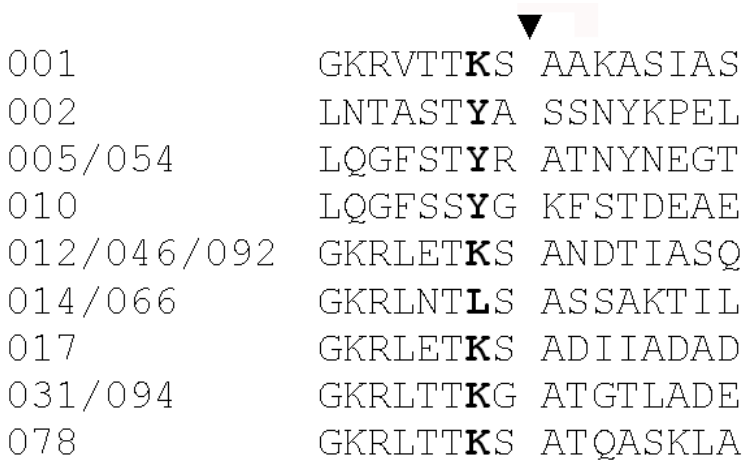


Figure 3.4.1 Alignment of predicted cleavage sites for trypsin or chymotrypsin within SlpA from fourteen *C. difficile* ribotypes (Eidhin *et al.* 2006).

Do changes in toxin production mean the bacterium can forgo an S-layer and still cause CDI? Or is immature SlpA still providing the rudimentary structural stability provided by an S-layer? To date, knockout of SlpA in *C. difficile* has been unsuccessful, indicating SlpA KO is lethal i.e. the resulting phenotype is unable to survive or that the ClosTron KO process does not account for the S-layer-less bacterium in its protocols e.g. ability to tolerate antibiotics or a certain nutritional requirements, creating the illusion of lethality. It is possible then that an S-layer negative strain of *C. difficile* may be created, perhaps by *slpA* complementation under expression control prior to *slpA* KO or by conditional KO, but using a different or modified KO process. It is possible for normally S-layer encompassed bacteria to live without an S-layer, as spontaneous mutant S-layer negative strains of *Camploybacter fetus* have been found (Tummuru & Blaser 1992) (although the regulator of S-layer is spontaneously mutated leaving the S-layer genes intact). Moreover, S-layer mutants appear to have decreased adherence to cell lines and decreased virulence in animal models (Janda *et al.* 1994; Sakakibara *et al.* 2007; Kern & Schneewind 2009). An S-layer negative strain of *C. difficile* is therefore likely to have impaired virulence, not only from the loss of adherence via its S-layer, or lack thereof, but through loss of accessory proteins directly or indirectly integrated with the S-layer.

3.4.7 Microscopy data

Despite there being no visible differences in CDΔCwp84 morphology compared to WT cells using Gram stain, a range of microscopy techniques were employed to determine if the presentation of immature SlpA on the cell surface of CDΔCwp84 affected cell surface ultra-structure.

3.4.7.1 Cryo-SEM

Freeze fracture cryo- SEM did not reveal any morphological differences between 24 or 48 hr solid phase cultures of WT or CDΔCwp84 respectively, however did find that cells of CDΔCwp84 are surrounded by large amounts extracellular material consistent with increased cell lysis and inter-cell adhesion.

Two layers were observed on the cell surface of both WT and CDΔCwp84, a smooth outer and an intermediate perforated layer. At present there are three known 'polymers' on the *C. difficile* cell surface: A surface carbohydrate coat (Ganeshapillai *et al.* 2008), the S-layer (Takumi *et al.* 1992; Cerquetti *et al.* 2000) and the peptidoglycan. The smooth outer layer may be one of two things: a polysaccharide capsule as found by Ganeshapillai *et al.* (2008) or the S-layer. Capsules are found in other S-layer containing bacteria e.g. *B. anthracis* (Mesnage *et al.* 1998). However, any structural relationship between the putative capsule and the S-layer of *C. difficile* requires further investigation. The ability to dissociate the smooth outer layer in sheet like fragments, similar to that seen in Figure 3.3.19, and those found by Takumi *et al.* (1991), together with embedded sectioning data suggests that the smooth outer layer is the S-layer. Replica freeze fracture TEM of bacterial S-layers seldom reveals holes in S-layer arrays (Bahl *et al.* 1997) suggesting the likely identification of the periodically perforated layer as the peptidoglycan. No cryo-SEM images of the peptidoglycan of Gram positive bacteria are available to confirm the predicted identification. The presence of pores in the peptidoglycan have been not previously described, casting doubt over this layers presumed identification, however these pores could be used to translocate S-layer subunits, for example, to and from the cell surface.

3.4.7.2 TEM

TEM analysis of embedded sections suggests that the LMW SLP portion of SlpA presented on the cell surface of CDΔCwp84 cannot lock into a regular array, or only forms a rudimentary array. This data supports the proposed orientation of the S-layer with the LMW SLP portion of SlpA outermost exposed to the environment linked, via a flexible (conserved cleavage) region, to the HMW SLP portion of SlpA, which can interact with the cell wall. However, TEM of CDΔCwp84 whole cells revealed that the Cwp84 mutant possesses a similar tetragonal lattice to that observed during freeze etching (Cerquetti *et al.* 2000), suggesting that immature SlpA is putatively able to pack on the cell surface. However, this interaction/packing may be temporary and easily disrupted or not 'locked' thus not affording the cellular, e.g. osmotic, protection afforded by an intact S-layer.

The presence of the thick layer underneath the peptidoglycan in CDΔCwp84 (Figure 3.3.21 B), compared to a very thin layer in WT cells, suggests that CDΔCwp84 is producing significantly more of this substance (presumed to be protein(s)) than the WT. The identification of this layer is unknown, but is likely to be as a result of (surface based) stress. It has been previously shown that environmental or chemical stresses change *C. difficile* surface protein expression (Deneve *et al.* 2008; Emerson *et al.* 2008) e.g. Cwp66 (Waligora *et al.* 2001), GroEL (Hennequin *et al.* 2001), putatively CD2767 and CD2797 (Dang *et al.* 2010). It is possible that the thickened layer (layer 3) is made up of surface protein species with N-terminal CWBDs e.g. CD2784 or Cwp66, as opposed to surface protein species with C-terminal CWBDS e.g. Cwp84 and SlpA. Moreover, an intriguing hypothesis is that this inner layer (layer 3) is made up of over-expressed, pre-peptidoglycan translocated SlpA which, coupled with an inability to lock SlpA on the cell surface, may be lost to the culture supernatant. Immunogold labelling studies using anti-SlpA antibodies may assist in testing this hypothesis.

3.4.7.3 AFM

AFM examination of 8 M urea extracted SLPs revealed that both WT and CDΔCwp84 SLPs formed raised plaque regions, putatively S-layer lattices, similar to those described by Chung *et al.* (2010). The hexagonal array of WT SLPs in AFM was significantly larger than the hexagonal array observed by Cerquetti *et al.* (2000) (using negative stained 8 M urea SLPs in TEM). Most significantly, the array formed by CDΔCwp84 SLPs displayed a completely different structure to the WT, and significantly larger than that observed *in vivo* on whole CDΔCwp84 cells (~10 nm). These data suggest 8 M urea extracts of CDΔCwp84 SLPs form a different array *in vitro* than *C. difficile* 630ΔErm SLPs.

Due to the limitations of AFM, it is impossible to infer which face of the putative S-layer array is being imaged. Cerquetti *et al.* (2000) suggests the *in vitro/in vivo* difference in S-layer arrangement is based on surface location, whereby *in vitro*, the large HMW subunit masks the thinner outer LMW subunit. The hexagonally packed intermediate (HPI) layer of *Deinococcus radiodurans* adsorbs strongly to freshly cleaved mica with the hydrophilic outer surface attached to the substrate, while the hydrophobic inner surface is exposed to the stylus (Muller *et al.* 1996). Given the proposed end-to-end structure of the HMW/LMW SLP complex (Fagan *et al.* 2009) the HMW may present itself as the uppermost surface, with the LMW attached to the substrate. The resulting array cannot be formed by CDΔcwp84 SLPs, suggesting that the immaturity of SlpA affects the ability of the S-layer to self-crystallise/self-assemble *in vitro*. Moreover, in order for CDΔcwp84 SLPs to form the tetragonal array, calcium was required in the wash buffer. The reason for this requirement is unknown but may be related to the ability of a mature S-layer array to retain chelated calcium required for array formation. Immature SlpA is unable to retain the calcium, thus for any array formation to occur calcium must be provided in the imaging buffer.

SLPs derived from glycine extracts from either WT or CDΔcwp84 did not appear to form an array that could be detected by AFM, which together with data suggesting that ultra-purification of the SLP subunits affects their ability to re-crystallise

(Takumi *et al.* 1991; Cerquetti *et al.* 2000), suggests an element(s) extracted in 8 M urea SLPs is necessary for array formation *in vitro*. However, H/L complex formation, a non-covalent interaction between the HMW and LMW SLPs, occurs regardless of this unknown (urea extractable) element and is calcium independent (Fagan *et al.* 2009).

3.4.8 Recombinant Cwp84

Cwp84 has been previously cloned and expressed in *E. coli* as both a GST-tag fusion (full and N-terminal cysteine protease domain) (Savariau-Lacomme *et al.* 2003; Janoir *et al.* 2004) and a full length His-tagged version (Janoir *et al.* 2007).

The key features of recombinant Cwp84 expression and purification from previous publications are therefore:

- A propensity to aggregate and form inclusion bodies
- Instability of protein and/or in its degrading abilities
- Expression of only the N-terminal cysteine protease domain may/may not have the same degrading activities as the full length enzyme
- Mature rCwp84 is approximately 61 kDa after 'automaturation' by incubation with DTT
- Cleavage of substrates can be inhibited by 100 µM E64 (cysteine protease inhibitor) and only occurs in reducing conditions

3.4.8.1 Expression of recombinant Cwp84

Expression of full length rCwp84 with a His_6-tag following Janoir *et al.*'s (2007) methodology expressed primarily insoluble. However, upon reducing the expression temperature to 16 °C, a soluble 58 kDa protein, presumed to be truncated rCwp84, with some enzymatic activity was purified.

The truncation may indicate that the cytoplasm of *E. coli* provides a sufficient reducing environment for rCwp84 to automature or *E. coli* proteases can cleave rCwp84 into a truncated form, which may or may not have been 'mature'. Prior

truncation, either by (auto)maturation or proteolysis, could explain the lack of affect of DTT on purified fractions of the 58 kDa rCwp84. Truncation i.e. removal of the CWBD(s) during expression in *E. coli* appeared to significantly decrease inclusion body formation and/or ease purification. However, the same 58 kDa species was purified regardless of N- or C-terminal His_6 tag creating uncertainty as to which end rCwp84 was being truncated from. The inclusion of other tags, e.g. NusA or GST, at the N-terminus of rCwp84 did not resolve truncation whereas material produced by *in vitro* transcription and translation, to avoid potential *E. coli* truncation, was found as an 84 kDa species.

Removal of the CWBD and expression of just the cysteine protease domain ($rCwp84_{33\text{-}497}$) was, as expected, more soluble due to decreased aggregation. N-terminal tag(s) of $rCwp84_{33\text{-}497}$ were removed indicating a similar auto-maturation or *E. coli* maturation/proteolysis seen for full length rCwp84. However, the size difference between putatively mature $Cwp84_{33\text{-}497}$ and immature $Cwp84_{33\text{-}497}$ is more consistent with the predicted MW loss upon removal of the cysteine protease pro-region (from 52 kDa to ~42 kDa) (Section 3.3.10). Together this data suggests that during recombinant expression of Cwp84, the protein may be truncated more than is estimated to be removed during (auto)maturation i.e. removal of the pro-region.

3.4.8.2 Cleavage of SlpA in CDΔ*cwp84* SLPs with recombinant Cwp84

SlpA in CDΔ*cwp84* SLPs did not appear to be cleaved into LMW and HMW SLPs, as seen for trypsin, with recombinant Cwp84. There are a number of explanations for this lack of cleavage:

- SlpA in CDΔ*cwp84* SLPs is presented differently *in vitro*
- rCwp84 is presented differently *in vitro*
- Combination of the two above.
- rCwp84 was not correctly folded/matured
- Trypsin is unique in its ability to access the site of cleavage
- Combination of all above points

It is likely that a combination of the above points explain the lack of SlpA cleavage by the rCwp84 produced in this study. It is interesting to note that Dang *et al.* (2010) co-expressed full length Cwp84, rather than just the cysteine protease domain, to confirm Cwp84 mediated cleavage of SlpA. This suggests the CWBDs of Cwp84 may be required spatial co-ordination between SlpA:Cwp84. Moreover, as only the conserved region of SlpA was expressed this may suggest that Cwp84 mediated cleavage of *full length* SlpA cannot be replicated *in vitro*.

The recombinant forms of Cwp84 produced in this chapter were able to cleave SlpA but not at the correct cleavage site and with poor efficiency (compared to other proteases tested on SlpA) but were inhibited by E64, suggesting cleavage was cysteine protease dependant activity. It is unknown which specific conditions *in vitro* e.g. pH, presence of divalent cations, detergents (Krupa & Mort 2000) or the presence of reducing environment, affect cleavage (or lack thereof).

Cwp84 is likely to be a highly efficient enzyme as culture medium supplemented with 1 mg/ml trypsin shows that 50% of SlpA remained uncleaved. Unfortunately, due to the amounts of recombinant Cwp84 produced, cell surface based cleavage of SlpA by recombinant Cwp84 was not determined. It is possible that surface exposed SlpA, is both accessible and oriented such that rCwp84, or other cysteine proteases e.g. papain, can cleave SlpA yielding WT–like SLPs.

3.4.9 Maturation of Cwp84

Janoir *et al.* (2007) suggests that recombinant Cwp84 produced in *E. coli* should 'self-mature' i.e. remove the proregion activating the cysteine protease, to give a 61 kDa protein *in vitro* upon exposure to DTT. However, both full length (putatively gylcosylated) 85 kDa and the 61 kDa versions are found on the cell surface.

The active cysteine at 116 aa in Cwp84 leaves only a maximum of 12 kDa (including signal peptide) that could be lost from the N-terminus before removal of C116. The proregion of papain is 107 aa (approximately 14.5 kDa) (Vernet *et al.* 1991; Vernet

et al. 1995). It is likely then, that the loss of 23 kDa (84 kDa to 62 kDa, as suggested by Janoir *et al.* (2007) means recombinant Cwp84 is not the same as the active form of Cwp84 found *in vivo*.

Intramolecular processing is found in a range of other cysteine proteases for example, after removal of the N-terminal pro-protein clostripain requires the autolytic removal of an internal nonapeptide to produce a (heavy and light chain) heterodimer (Dargatz *et al.* 1993). Furthermore, gingipains (C25 family cysteine protease) undergo a complex scheme resulting in a non-covalent multidomain, multifunctional complex anchored into the outer membrane by a glycated, C-terminal heamaglutanin/adhesin domain (Potempa *et al.* 2003).

3.4.10 Cwp84:SlpA hypotheses

3.4.10.1 SlpA:Cwp84 genetic control

slpA and *cwp84* are upregulated during exposure to sub-MIC levels of antibiotics (Deneve *et al.* 2009a). This co-ordinated upregulation, particularly of *cwp84*, highlights how during stress the bacterium acts to increase the structural consistency of the protective S-layer and/or change its surface composition. The decreased cell integrity of CDΔCwp84 may also result in upregulation of *slpA*, similar to exposure to oxygen or other detrimental environmental stimulus (Deneve *et al.* 2008; Emerson *et al.* 2008).

S-layer mRNA stability is thought to be high at around 11-22minutes (Chu *et al.* 1993; Bahl *et al.* 1997; Kahala *et al.* 1997) (compared to other prokaryotic transcripts half-lives of minutes even seconds). It is therefore likely that *C. difficile slpA* mRNA retains this long half-life feature i.e. minimal upregulation required to increase protein levels. The higher upregulation of *cwp84* may suggest a short half-life for *cwp84* mRNA. In *Thermus thermophilus* HB, SlpA (100 kDa main S-layer) and SlpM (52kDa alternative S-layer like protein) is controlled via a regulated circuit involving SlpA which regulates its own expression (and SlpM), together with SlrA which acts as a repressor, and SlpM as an activator (Fernandez-Herrero *et al.* 1997).

Does the apparently co-ordinated expression of *slpA*:*cwp84* therefore point to a regulatory mechanism? For example *cwp84* or Cwp84 directly regulating *slpA* or *slpA*/SlpA regulating its own expression, deregulated by Cwp84 inhibition.

Control of Cwp84 could occur via the action of its pro-peptide attenuating already active Cwp84, or acting on alternative pathways (Maubach *et al.* 1997; Yamamoto *et al.* 2002; Burden *et al.* 2008). Control of SlpA cleavage, potentially *slpA* transcription itself, may be performed by up or down regulating *cwp84* transcription and translation (and activation). As Dang *et al.* (2010) suggests expression of *slpA* is increased in response to chemical inhibition of Cwp84, an interesting extension to Dang *et al.* (2010)'s work could be investigating the inhibition of Cwp84 by its own propeptide and its effect on *slpA* expression.

3.4.10.2 Cwp84 mediated cleavage of SlpA

There are two possibilities regarding the maturation of the SlpA by Cwp84 in *C. difficile* :

- Direct cleavage I- Cwp84 alone matures SlpA
- Direct cleavage II– Cwp84 cleaves SlpA but requires co-factors
- Indirect cleavage - Knockout or inhibition of Cwp84 prevents something else cleaving SlpA, or prevents SlpA from being cleaved by something else.

The data presented in this chapter, and the work of Dang *et al.* (2010) provides convincing evidence that Cwp84 alone matures SlpA.

However, Dang *et al*. (2010) demonstrated cleavage of an SlpA fragment but not cleavage of the full SlpA precursor. The co-purification of Cwp66, CD2797 and CD2767 (Dang *et al.* 2010) suggests other proteins may be involved in with S-layer biogenesis, including its cleavage. Perhaps, SlpA is cleaved by Cwp84 but is then bound (non-covalently) to the underlying cell wall e.g. peptidoglycan or SWCP, by these 'accessory' proteins. S-layers with SLH domains are non-covalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation (Mesnage *et al.* 2000). The highly conserved, positively charged N-

terminal region of the *Geobacillus stearothermophilus* S-layer protein, SbsC (which like the *C. difficile* S-layer does not have SLH domain(s)) binds to a SCWP of negatively charged N-acetyl glucosamine, glucose, and 2,3-dideoxy-diacetamido mannosamine uronic acid in the molar ratio of 1:1:2 (Pavkov *et al.* 2008). Furthermore, the N-terminus of the *Bacillus stearothermophilus* PV72/p2 S-layer binds to a SCWP incorporated into the peptidoglycan (Ries *et al.* 1997). The possibility of a SCWP removable by urea, and not glycine (low pH), adds credence to the argument that urea SLP extracts contain an element allowing correct/full S-layer self assembly *in vitro* as demonstrated by AFM. However, the large amounts of SWCP that would be required, up to 20% of the cell surface (Ries *et al.* 1997), have not been found to date in *C. difficile*. Although, *C. difficile* has been shown to express two highly complex cell-surface teichoic-acid-like polysaccharides (PS-I and PS-II), one was composed of a branched pentaglycosyl phosphate repeating unit and the other was composed of a hexaglycosyl phosphate repeating units (Ganeshapillai *et al.* 2008).

It is therefore more likely, that Cwp84 cleaves SlpA but correct S-layer lattice assembly may be assisted by co-factors.

3.4.10.3 Other roles for Cwp84

A direct role for Cwp84 as a colonisation factor through the digestion of extracellular matrixes has been proposed. While the findings in this chapter are not inconsistent with Cwp84 having additional roles involved in matrix degradation, they do demonstrate that the enzyme is not essential for CDI in the animal model which could be due to redundancy of function within the surface-associated components.

3.4.10.4 Location of Cwp84 and location of cleavage.

Convincing evidence suggests that Cwp84 cleaves SlpA, however where on/in the cell (envelope) does this cleavage occur?

Immature SlpA found on the surface in CDΔCwp84 lacks its signal peptide, signal peptidase cleavage therefore precedes cleavage into SLP subunits i.e. post

translocation across the plasma membrane. The S-layer of *Haloferax volcanii* is also post-translationally modified following translocation across the plasma membrane (Eichler 2001). Cleavage of the S-layer in *C. difficile* is likely either underneath, amongst or on top of the peptidoglycan (in the S-layer) (or a combination of all three).

During early work on the S-layers of *C. thermohydrosulfuricum* and *C. thermosaccharolyticum,* a surplus of S-layer subunits was observed in the region between the cytoplasmic membrane and the cell wall (Sleytr & Glauert 1976). Furthermore, Breitwieser *et al.* (1992) suggests an additional S-layer on the inner surface of the peptidoglycan and goes on to suggest that this profile is characteristic for cell wall preparations of many S-layer-carrying Gram-positive bacteria. It is possible then that the location of the CWBD (at the N- or C-terminus) determines the arrangement of proteins beyond the cell membrane, potentially either side of the peptidoglycan. The location of the CWBDs in SlpA and Cwp84 would present both in a similar manner on the cell surface. It is also likely that Cwp84 is activated at or near the point of cleavage, as activation before that may result in unwanted proteolysis. The insertion sites of new SlpA precursor would be the most obvious locus for Cwp84 (and Cwp84 mediated cleavage).

3.4.11 Conclusion

Cwp84 is a surface located cysteine protease which matures the S-layer of *C. difficile*. Cleavage of the S-layer is not a pre-requisite for virulence and inhibition of S-layer cleavage may induce putative stress responses increasing toxin production. Based on the data presented, targeting Cwp84 to decrease S-layer integrity may allow other pharmacological treatments. However, the inadvertent changes in cell metabolism to counteract surface based changes are unlikely to decrease CDI pathology and may cause increased mucosal damage. Therefore, inhibition of S-layer biosynthesis may not be an immediate route for therapeutic investigation. Adhesion to the gut mucosa or prevention of spore formation to prevent spread and potentially relapse may be other surface based alternatives.

CHAPTER 4 –
Immunotherapy of *C. difficile*

4 CHAPTER 4 – Immunotherapy of *C. difficile*

4.1 Introduction

4.1.1 *C. difficile* treatment- current options

The current treatment for *C. difficile* infection (CDI), beyond removal of the offending broad-spectrum antibiotic, is the administration of metronidazole then, if unsuccessful, treatment with vancomycin.

Vancomycin is particularly suitable as a gastrointestinal antibiotic as it is completely restricted to the colonic lumen i.e. not absorbed, and found at levels more than 100 times higher than the highest minimum inhibitory concentration reported (Bartlett 2008a). However, Vancomycin has several problems, the first is the cost (approximately £120 for a 10-14 day course, 125 mg given orally four times daily) and secondly vancomycin promotes faecal colonisation with vancomycin-resistant enterococci (Gerding 1997; Al-Nassir *et al.* 2008). However, the rates of relapse for both vancomycin and metronidazole are approximately 15%–25% (Bartlett 2008a). Antibiotics thus cause *and* treat CDI, however treatment failure means some patients suffer repeated rounds of CDI (aka relapses) (McFarland 2005). Relapse (or recurrence) is likely to result from either continued exposure to *C. difficile* spores from the environment (healthcare or domestic) or from an endogenous source (Pepin *et al.* 2006).

Therefore, there is a need to develop new therapeutic strategies which both cure the initial CDI episode and prevent relapses. A range of 'new' antibiotics is being evaluated to treat CDI, each with its own set of advantages and disadvantages (Bartlett 2009; Gerding & Johnson 2010). Interest in non-antibiotic therapies has increased as they are likely to leave the protective gut flora intact (or allow it to recover) (Borriello 1990), a facet likely to be key in decreasing relapses. Non-antibiotic therapies fall into two categories based on preventing or treating different parts of the CDI infection cycle:

- Competitive inhibition- by re-establishment (or re-insertion) of the normal colonic flora (Bakken 2009), insertion of inhibitory species (Tung *et al.* 2009) or insertion of non-toxinogenic *C. difficile* (Borriello & Barclay 1985).
- Preventing damage caused by the *C. difficile* toxins- using chemicals or resins to bind the toxin e.g. Tolevamer (Weiss 2009) or by neutralisation of the toxicity e.g. using immunoglobulin (O'Horo & Safdar 2009).

4.1.2 The immune system and *C. difficile* infection

A large majority of the population have circulating antibodies to *C. difficile* toxins (Viscidi *et al.* 1983) making detection and/or discrimination of any CDI induced specific antibody responses (serum or mucosal) difficult. Furthermore, the extent of and the protective nature of any generated immune responses is debated due to a lack of understanding in the pathogenesis of CDI. In particular, which of the *C. difficile* toxins are required for disease (Lyras *et al.* 2009; Kuehne *et al.* 2010) and thus to which toxin(s) an immune response is generated, what kind of response is generated (in terms of antibody type) and whether these responses (or a lack of) are protective. Collectively, data appears to suggest that progression of CDI is connected with humoral immune deficiencies in response to *C. difficile* infection (Aronsson *et al.* 1985; Warny *et al.* 1994; Johnson 1997; Kyne *et al.* 2000; Kyne *et al.* 2001; Katchar *et al.* 2007) (Summarised in Figure 4.1.1).

CDI is also characterised by induction of an inflammatory cascade resulting in damage from severe inflammation. The extreme inflammatory reaction is thought to be derived from *C. difficile* toxin activation of cell mediated immunity causing release of pro-inflammatory cytokines including interleukin-8 (IL-8), IL-1β and TNF-α (Pothoulakis *et al.* 1988; Linevsky *et al.* 1997; Ng *et al.* 2010).

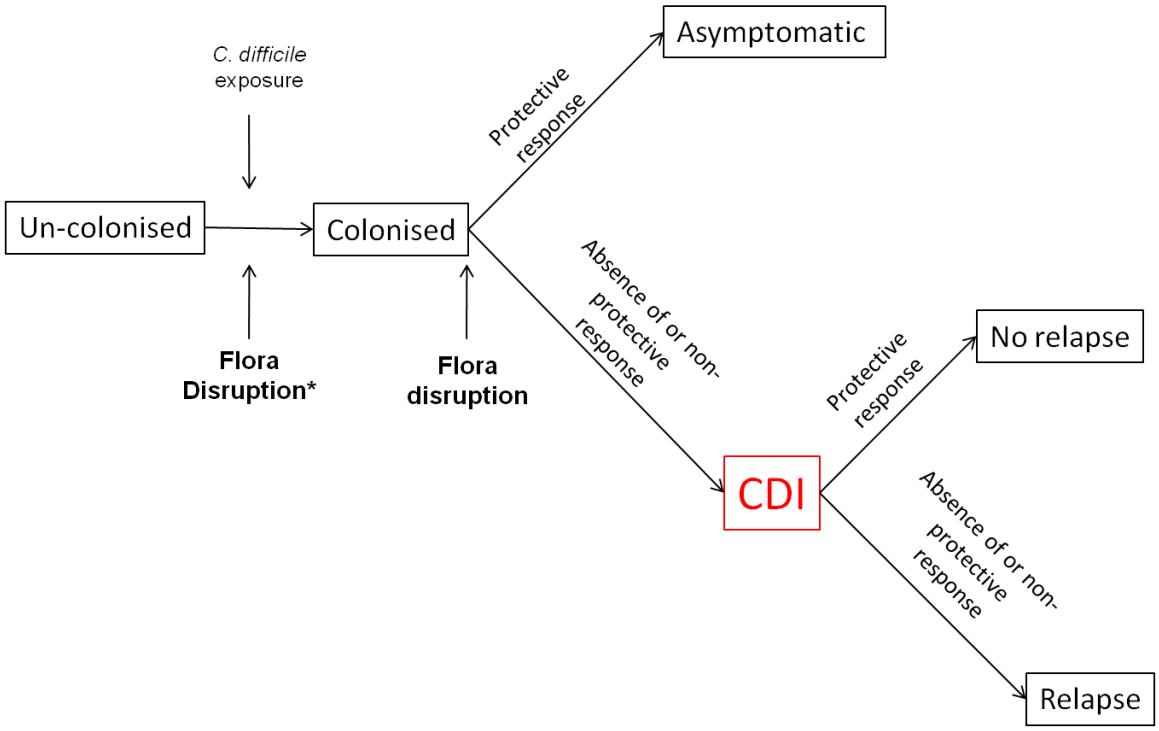


Figure 4.1.1 Development of *C. difficile* infection. Figure adapted from Wilcox & Minton (2001). **C. difficile* may be acquired without flora disruption.

Taken together, if one can increase the systemic humoral immune response-protection from direct toxin or inflammatory damage, may be possible i.e. a mucosal response may not be necessary. This proposal is supported by Giannasca *et al.* (1999), who found that a protective response in hamsters, afforded by high levels of toxin neutralising antibodies in the serum, was only obtained with rectal and intramuscular immunisation of *C. difficile* toxin toxoids and not by other mucosal routes.

4.1.2.1 *C. difficile* toxin immunotherapy

C. difficile immunotherapy seeks to increase the immune response via active or passive immunisation and given that the toxins alone can replicate many hallmarks of CDI (Borriello & Barclay 1985), the toxins represent a key target for vaccine development. Active vaccination using toxoids of toxins A or B is attractive and has proven to be successful in limited trials (Kotloff *et al.* 2001; Aboudola *et al.* 2003; Sougioultzis *et al.* 2005) but the ability of immunocompromised patients to respond remains a key unanswered question.

Early studies suggested that colostrum was able to neutralise *C. difficile* toxins, thus providing (passive) protection to neonates of which a high proportion are colonised with *C. difficile* (Wada *et al.* 1980; Kim *et al.* 1984). Work by Giannasca *et al.* (1999) also found passively administered antibodies, to toxoids of toxins A and B, were protective in the hamster model. Since then, a number of studies have assessed the protective effects of *in vitro* administered immunoglobulin (IVIG) in humans, to one or both *C. difficile* toxins (Leung *et al.* 1991; Lyerly *et al.* 1991; Salcedo *et al.* 1997; Wilcox 2004; Juang *et al.* 2007) although limited trial sizes, no set dose ranges and mixed results means further research is required (O'Horo & Safdar 2009; Abougergi *et al.* 2010). To date, full protection appears to only be provided with an anti-toxin A *and* B preparation (Kink & Williams 1998; Babcock *et al.* 2006; Lowy *et al.* 2010) confounding the role of each toxin in CDI.

Antibodies (naturally raised or administered) to toxins do not appear to have any effect on colonisation, as *C. difficile* can be cultured from asymptomatic carriers (who have a protective immune response) (Shim *et al.* 1998; Sanchez-Hurtado *et al.* 2008). Moreover, the increased antibody response to *C. difficile* toxins seen by Kyne *et al.* (2000) in asymptomatic individuals, occurred *after* colonisation.

4.1.3 The role of *C. difficile* surface proteins and their associated immune responses in CDI

The surface proteins of *C. difficile* are likely to play a key role in colonisation and subsequent infection, as a result the surface proteins of *C. difficile* are potential targets for immune attack, in particular the S-layer as the major surface antigen of the bacterium. Due to their significant surface coverage and key role in adhesion/colonisation, S-layers often undergo antigenic variation via gene duplication, switching and recombination events, to try evade the immune response e.g. *Campylobacter fetus* varies the expression of its SLPs via genomic rearrangements of S-layer genes *in vivo* during infection (Sara & Sleytr 2000; Thompson 2002). Despite this variation, S-layers induce immune responses, as serum from infected patients often reacts with extracted S-layers (Janda *et al.* 1994). Moreover, the serum IgG response to *Bacteroides forsythus* S-layer appears to be lower in healthy controls than in cases with adult and early-onset

periodontitis (Yoneda *et al.* 2003), suggesting that an inability to mount a sufficient immune response to surface based antigens may be associated with disease progression.

Upon blotting surface extracts of *C. difficile* with CDI patient serum, Pantosti *et al.* (1989) demonstrated that *C. difficile* non-toxin proteins elicited an immune response, specifically the proteins of the S-layer and its constituents. As expected, based on surface coverage, the majority of the humoral antibody response appears to be directed to the main S-layer (Drudy *et al.* 2004; Wright *et al.* 2008). However, other 'minor' uncharacterised proteins are also capable of generating an immune response in the same number of patients as the SLPs, particularly the uncharacterised S-layer paralog CD2791 (Cwp2) (Wright *et al.* 2008). As with many proteins certain regions of *C. difficile* surface proteins appear better epitopes than others, for example, the C-terminal region of the putative heat shock adhesin Cwp66 raises a greater response than the N-terminal (CWBD) region (Pechine *et al.* 2005a; Pechine *et al.* 2005b). Antibody responses to Cwp66, the flagella core and cap (FliC and FliD) and a fibronectin binding protein (Fbp68) also appeared to be greater in controls than CDI cases, suggesting an immune response to surface proteins, other than the S-layer, could play a role in a host defence mechanism (Pechine *et al.* 2005a). Rectal immunisation using a formalin killed whole-cell preparation (which would have contained all the 'normal' *in vitro* vegetative cell surface proteins) in conjunction with toxoid culture filtrate appeared to be able to induce an, albeit non-protective, immune response (Torres *et al.* 1995).

4.1.3.1 Immuno-modulation of S-layer proteins

C. difficile surface proteins, like its toxins, putatively have the ability to modify the immune system beyond creation of an antibody response. Ni Eidhin *et al.* (2008) suggested that the S-layer of *C. difficile* may divert the immune response in favour of a co-administered antigen (a feature utilised in the generation of vaccines by genetic fusion of antigens with S-layer proteins e.g. the major birch pollen allergen Bet v 1 fused with the S-layer from *Geobacillus stearothermophilus* ATCC 12980 (Gerstmayr *et al.* 2007) or avian flu nucleoprotein on *Bacillus thuringiensis* S-layer

(Liu *et al.* 2008b)). In support of this hypothesis, Sanchez-Hurtado & Poxton (2008) found that *C. difficile* SLP extracts appeared to increase the toxicity of toxin A when administered to Vero and Caco2 cells *in vitro,* while Brun *et al.* (2008) found the LMW SLP displayed adjuvant-like properties to a co-administered antigen in mice immunised via the mucosal intranasal or subcutaneous route.

Another hypothesis is that *C. difficile* SLPs are able to induce the release of elevated amounts of IL-1β and IL-6 pro-inflammatory cytokines by resting monocytes, similar to those induced by LPS from *E. coli* (Ausiello *et al.* 2006), a result similar to the modulation of dendritic cells and T cell function by the S-layer of *Lactobacillus acidophilus* NCFM, as demonstrated by S-layer knockout (Konstantinov *et al.* 2008). Although recent research by Sekot *et al.* (2011), found that an S-layer negative strain of *Tannerella forsythia* caused significantly higher levels of pro-inflammatory IL-1β, TNF-α, and IL-8 compared with wild-type, suggesting the S-layer may even provide an immune evasive function. The immune responses of CDI patients to surface proteins, yet their apparently non-protective nature, could point to a *C. difficile* (surface based) immune evasion system i.e. generation of high titre but non-protective antibodies or that the patients themselves cannot raise an adequate, protective (anti-surface protein) immune response.

4.1.3.2 *C. difficile* surface protein immunotherapy

To assess if an immune response to *C. difficile* surface proteins provides any protective effects against CDI, several studies have investigated active and passive immunisation of hamsters and the ability of immune response(s) to specific surface protein(s) to decrease *C. difficile* CDI and colonisation.

The ability of an S-layer extract to elicit a protective immune response by active immunisation was investigated by Ni Eidhin *et al.* (2008) using *C. difficile* R13537 (Ribotype 001). Purified LMW & HMW SLP were able to elicit a good immune serum IgG response particularly when administered at >25 μg with alum intraperitoneally (i.p.) in hamsters, although, this response was non-protective. Moreover, SLPs administered with Ribi (monophosphoryl lipid A and synthetic trehalose

dicorynomycolate), with or without Cholera toxin, did not appear to protect challenged hamsters or display any correlation with antibody responses.

Active immunisation with specific surface proteins (rather than a purified S-layer extract) appeared to be able to decrease colonisation > 6 days post challenge with *C. difficile* 79–685 in the mouse model (Pechine *et al.* 2007). In particular, rectal administration of encapsulated recombinant flagella cap protein (FliD) in combination with a flagella preparation or recombinant Cwp84 (plus cholera toxin), although variable, appeared to decrease colonisation statistically significantly more than PBS over time. Moreover, a rectally administered cell wall extract (produced by vortexing vegetative cells with glass beads) was particularly successful at decreasing colonisation from day 13, suggesting a multivalent surface protein preparation may be the most successful at CDI prevention. Assessment of immunisation route suggested that the rectal administration was deemed to be most effective at generating a serum IgG *and* mucosal IgA (from intestinal lavage), while serum IgG was highest from subcutaneous immunisation. Which of the immunoglobulin species provided colonisation resistance or the extent of any disease progression was not alluded to, but perhaps highlights the differences between hamster (time-to-death) and (chronic) mouse models (Chen *et al.* 2008).

Passively administered anti-SLP antibodies appeared to extend the survival of hamsters challenged with *C. difficile* R13537 (ribotype 001) (clindamycin resistant) vegetative cells (pre-incubated with the aforementioned antiserum). Whereby, orogastrically anti-SLP antibodies administered 7 hr before challenge then 6, 17 and 24 hr after challenge, ultimately failed to protect hamsters which still developed diarrhoea within 2-5 days post challenge and eventually died (O'Brien *et al.* 2005). Although, O' Brien *et al.* (2005) suggested that anti-SLP antibodies exert their protective effect by increasing phagocytosis of *C. difficile* cells, as demonstrated in differentiated THP-1 monocytes.

The surface proteins of *C. difficile* are thus able to generate an immune response which may decrease colonisation and prevent (or prolong survival) CDI in the hamster model.

The aim of this study was therefore to investigate the surface proteins of the *C. difficile* Stoke Mandeville (R20291) 'hypervirulent' ribotype 027 strain, in particular to examine if passively administered high titre anti-SLP antibodies were able to prevent CDI in the hamster model. Presented here are data suggesting that the SLPs of the 027 ribotype are largely conserved and while anti-SLP antibodies prevent binding to cell lines, they do not prevent CDI when administered systemically and thus their usefulness in a clinical setting maybe limited.

4.2 Chapter 4 specific methods

4.2.1 *C. difficile* SLP extract Toxin A contamination ELISA

C. difficile R20291 and VPI10463 SLP extracts were assayed by sandwich ELISA as described in General Methods section 2.7.1.

4.2.2 Polyclonal anti-SLP serum production

4.2.2.1 Small scale in guinea pigs

C. difficile R20291 and VPI10463 SLP extracts (20 µg diluted in 50 mM HEPES plus 0.15 M NaCl pH 7.4) were administered, with Titremax Gold™ adjuvant, subcutaneously. Initial immunisation was followed by 2 further boosts at approximately 6 and 10 weeks. Guinea pigs had test bleeds at 7 weeks and were fully bled at 11 weeks.

4.2.2.2 Large scale production in sheep

C. difficile R20291 and ARL001 SLP extracts (100 µg diluted in PBS) were administered, with Freund's incomplete adjuvant, subcutaneously. Initial immunisation was followed by 3 further boosts at approximately 4, 8 and 12 weeks. Sheep were bled at week 14.

4.2.2.3 Anti-SLP serum titre estimation

Serum titres were estimated by indirect ELISA based on Drudy *et al.* (2004) except using 10 µg/ml *C. difficile* SLP extract, PBS/T plus 5% foetal calf serum as blocking buffer and appropriate anti-species HRP antibody (diluted to 1:2000 with blocking buffer).

4.2.2.4 Large scale purification of anti- R20291 SLP IgG

Ovine polyclonal anti-SLP antibodies were purified from sheep serum using an adapted caprylic acid precipitation method (Steinbuch & Audran 1969) (Abdulla, I., personal communication (MicroPharm Ltd.)). Serum was diluted to twice volume with 0.9%w/v sodium chloride followed by addition of caprylic acid (octanoic acid) to a final concentration of 2% w/w. The mixture was then stirred vigorously for 30 min, left to stand for further 60 min, then centrifuged at 3488 g for 60 min. The supernatant was then collected, filtered through 0.45 µm filter and dialysed against

50 mM HEPES plus 0.15 M NaCl pH 7.4. Antibodies were then concentrated to ~50 mg/ml (A_{280}) using a 10 kDa MWCO Vivapsin 20 spin concentrator and filtered through 0.2 µm filter and stored at 2-8 °C (short term) or at -20 °C (long term).

4.2.3 Immunofluorescent binding of anti-SLP antibodies to *C. difficile*

Purified guinea pig anti-R20291 SLP IgG was labelled with Alexa Fluor® 488 dye using the Alexa Fluor® 488 protein labeling kit (Molecular Probes, USA) and was then assessed for its binding to *C. difficile* cells. *C. difficile* whole cells at OD_{600} ~0.7, were harvested then re-suspended in PBS and 10 µl spread onto a glass slide. The cells briefly were stained with ethidium bromide (1:10,000 dilution in PBS) for 10-15 sec, then rinsed with PBS to aid visualisation. Slides were then incubated with 10 µg/ml Alexa 488 labelled guinea pig anti-SLP IgG for 30 min, followed by washing.

For competition assay, slides were first blocked with 3% w/v BSA in PBS then incubated with unlabelled guinea pig anti-SLP IgG (100 µg/ml) for 1 hr, followed by incubation with labelled guinea pig anti-SLP IgG (10 µg/ml) for 30 min with a PBS rinse in between. All incubations were at room temperature at 100% relative humidity and all specimens were examined with 100x oil-immersion objective using a Nikon fluorescent microscope with the FITC filter.

4.2.4 Bacti-ELISA

Binding of anti-SLP IgG to *C. difficile* whole cells was analysed by indirect colorimetric ELISA based on Elder *et al.* (1982) and Prieto *et al.* (2003). Maxisorb microtitre plates (Nunc) were coated for 1 hr at 37 °C with 100 µl of washed stationary phase *C. difficile* cell suspension (approx. OD_{600} = 1.4) or 10 µg/ml SLPs (positive control) diluted in PBS. Plates were blocked with 3% BSA in PBS/T (blocking buffer) for 1 hr at 37 °C, followed by washing. Plate(s) were then incubated for 1 hr at 37 °C with 100 µg/ml guinea pig anti-SLP IgG or 1:100 dilution of *C. difficile* toxin B2 fragment antiserum (to assess *C. difficile*'s generic IgG binding ability). After washing, plates were incubated with goat anti-guinea pig IgG HRP conjugate (Sigma) (1:2000) for 1 hr at 37 °C. The plate(s) were washed a final time

and developed with TMB substrate and TMB stop solution then read at 450 nm using a Tecan™ Sunrise microtitre plate reader. All washes were performed 3 times with PBS/T.

4.2.5 Sequencing of slpA from a range of 027 strains

slpA was amplified using primers from Qazi *et al.* (2009) using stationary phase *C. difficile* gDNA extracted (as per General Methods Section 2.1.11) from the following *C. difficile* strains (Table 4.2.1):

Table 4.2.1 *C. difficile* PCR ribotype 027 strains used in this chapter

Code	Isolation location	Approximate Isolation date
KCH#2 9706	Kings College Hosp isolate	Feb 2006
918186A	Poole	27/9/07
918692B	Southampton (1)	3/11/07
918341E	Medway	6/9/07
918737F	Eastbourne	19/10/07
918380U	Portsmouth	4/10/07
918443W	Crawley	4/7/07
CDRNE 027 Reference strain		from EQA panel distributed 12/07
R20291	027	Anaerobe reference laboratory (ARL) 027 –Stoke Mandeville isolate.

All 027 strains were a kind gift from Steve Green (HPA Regional Office, Southampton)

4.2.6 Adhesion assay

The *in vitro* adhesion assay was developed by L. Blandford. Caco-2 cells were seeded into 8 well Lab-Tek™ chamber slides at a 1:2 dilution in MEME (approx 5.69 x 10^4 cells/well) for 14 days prior to the assay to ensure full differentiation. 18 hr *C. difficile* culture was harvested by centrifugation then washed twice with an equal volume of Dulbecco's phosphate buffered saline (DPBS), before re-suspension in

MEME ± ovine anti-R20291 SLP IgG (in 50 mM HEPES 0.15 M NaCl pH 7.4) or ovine anti-R20291 crude spore serum. Caco-2 cells in chamber slides were washed twice with 0.4 ml per well of DPBS before addition of 0.4 ml of bacterial suspension to each well. Slides were incubated for 2 hr 20 min at 37 °C in anaerobic conditions followed by washing three times with DPBS. Slides were then fixed with methanol at -20 °C for 5-10 minutes and stained with Giemsa stain (Sigma, Poole, UK) diluted with Wright-Giemsa stain phosphate buffer pH 6.8 (Park Scientific Ltd, Northampton, UK) before microscopic observation at x1000 magnification using an oil immersion lens. Enumeration was performed using a single-blind experimental design.

4.2.7 Passive immunisation of hamsters with anti-SLP antibodies

Syrian hamsters (80-100 g) were housed (2 per cage) in isolator cages, fitted with air filters on lids to minimise contamination between groups. Groups of 24 hamsters were divided into two 'Test' subgroups of 10 and a 'Control' subgroup of 4 animals. All 24 hamsters were weighed and administered clindamycin (2 mg in 0.4 ml sterile H_2O) by the orogastric route on Day 0. In the morning of Day 4, hamsters in both 'Test' subgroups were challenged with 2.0 x 10^5 colony forming units of *C. difficile* spores in 0.4 ml DMEM, given orogastically. In the afternoon of Day 4, 1 ml of purified ovine anti- *C. difficile* R20291 SLP IgG (approx 500 mg/kg) was administered intraperitoneally. On days 7 and 10, 1 ml of purified sheep anti- *C. difficile* R20291 SLP IgG was administered intraperitoneally as before. All animals were weighed daily and monitored 6 times/day, up to Day 24, for disease symptoms which included: diarrhoea, weight loss, lethargy and tender abdomen (Sambol *et al.* 2001; Babcock *et al.* 2006). Hamsters were scored on a 0-3 scale for disease symptoms, animals in advanced stages of disease or became immobile were euthanised.

4.3 Results

4.3.1 Extraction of surface layer proteins from a range of *C. difficile* strains

It has previously been suggested that the surface of epidemic (hypervirulent) strain(s) of *C. difficile* may differ from that of non-hypervirulent strains resulting in increased adherence to human intestinal epithelial cells (Cartman *et al.* 2010). To investigate this, the low pH extracted surface proteins of the putatively 'hypervirulent' ribotype 027 were compared to high toxin producer VPI10463, low toxin producer 630 (although using the erythromycin resistant mutant described in Hussain *et al.* (2005)) and the clinically relevant 001 ribotype (Brazier 1998; Cheknis *et al.* 2009).

For all *C. difficile* strains analysed, there are two main (S-layer) bands which vary between strains, except for VPI10463 and ARL001 (See Table 4.3.1).

Table 4.3.1 Variation of the main S-layer bands of various *C. difficile* strains as measured by 1D SDS-PAGE.

Strain	LMW SLP (kDa)	HMW SLP (kDa)
R20291	40	50
630ΔErm	40	47
VPI10463	39	49
ARL001	39	49

Protein bands >50 kDa vary in intensity amongst the strains tested. A consistent band at 66 kDa is seen in all strains, while the extract of *C. difficile* 630ΔErm contains a further band at 120 kDa not visible in other extracts (Figure 4.3.1).

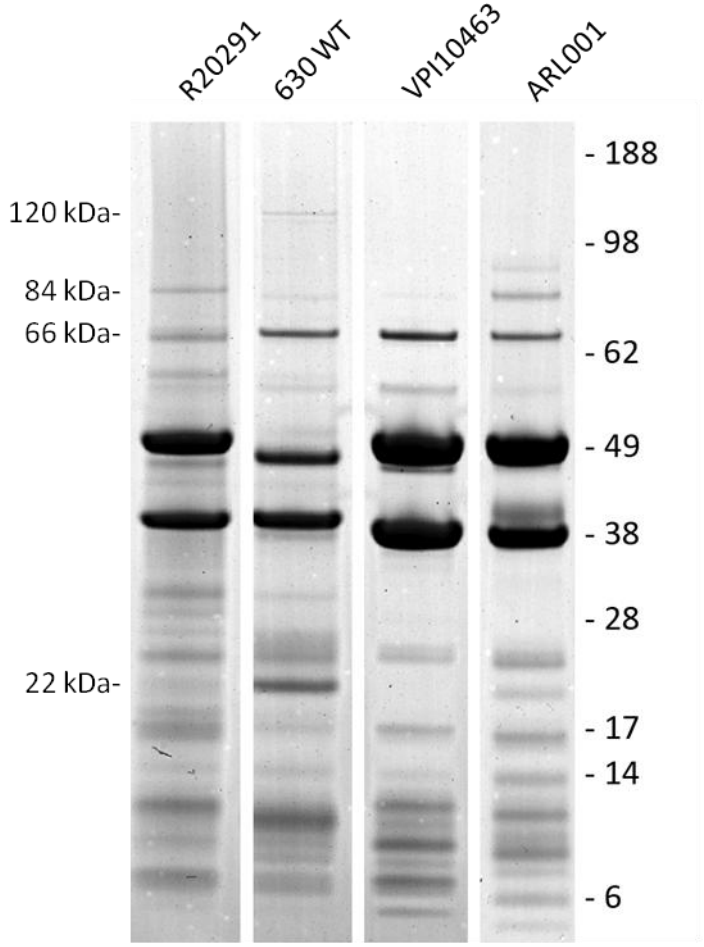


Figure 4.3.1 Comparison of SLP extracts from hypervirulent R20291 with *C. difficile* 630ΔErm, VPI10463 and ARL001 by 1D SDS-PAGE.

Bands at 120 kDa, 84 kDa, 66 kDa and 22 kDa in the SLPs of *C. difficile* 630ΔErm were putatively identified in other sections of this thesis as the following:

- 120 kDa band identified as the C-terminal 120 kDa mature fragment of cell surface protein (putative haemagglutinin/adhesin) CwpV (CD0514) - predicted MW of 160 kDa (Emerson *et al.* 2009a) (Chapter 3 Section 3.3.4.1)
- 84 kDa band- putatively the band identified as the S-layer precursor protein SlpA (*slpA* (*cd2793*) - predicted MW of 76.1 kDa (Chapter 3 Section 3.3.4)
- 66 kDa band - identified as the cell surface protein (putative S-layer protein precursor) CD2791 a.k.a. Cwp2 (*cd2791*) - predicted MW of 66.4 kDa (Chapter 5 Section 5.3.3.3).

- 22 kDa band – identified by N-terminal sequencing (MNLKGT) as rubrerythrin (energy production and conversion protein) - predicted MW of 20.5 kDa.

Multiple additional bands *below* the two main S-layer bands are seen and display a degree of strain variation (Figure 4.3.1). Interestingly, an SLP extract from 24 hr culture *C. difficile* VPI10463 extracted only the LMW S-layer and not both SLP species, supporting the proposed spatial arrangement with the LMW S-layer outermost and thus easier to extract.

There may be protein differences in the 027 ribotype SLPs (R20291 reference strain in this study) compared to other *C. difficile* strain surface extracts, however due to the insufficient resolving power of 1D SDS-PAGE this study cannot comment on such differences. There does, however, appear to be visible differences between the high (VPI10463) and low (630) toxin producers, the visible presence of a 120 kDa (CwpV) species in 630 and low MW (<35 kDa) fragmentation pattern.

4.3.2 Sequencing of *slpA* from a range of 027 Strains

The entire S-layer precursor gene *slpA* (*cd2793*) was amplified from gDNA extracted from a range of regional *C. difficile* 027 ribotype isolates and sequenced to ascertain any sequence variability or conserved domains *within* the 'epidemic' 027 ribotype. The *slpA* gene sequence of the nine isolates was 100% identical. Given the proposal of using *slpA* sequencing as a typing system (Karjalainen *et al.* 2002; Kato *et al.* 2005; Kato *et al.* 2009), a high degree of similarity is to be expected. This indicates that certainly for regional isolates, the largest surface antigen (main S-layer) is exactly the same.

BLAST searches of the following published *C. difficile* 027 ribotype genome sequences, revealed that *slpA* of R20291, 196, 855 and QCD-32g58 are 100% identical with each other, and 100% identical with the *slpA* sequences obtained from the 027 ribotype isolates in this study. *slpA* of *C. difficile* strain BI1 differs by one base pair (G→A 1811 bp) resulting in a P156L change.

- R20291 a.k.a. SM (Stoke Mandeville) UK, 2006
- 196 - France, 1985
- BI1 - USA, 1988
- 855 - USA, 2007
- QCD-32g58 Quebec -Canada, 2005

A highly conserved S-layer amongst the 027 ribotype could suggest that it is a highly evolved gene i.e. is more adherent or is less available to immune recognition, and therefore worth preserving.

4.3.2.1 *In silico* analysis of other cell wall binding domain containing surface protein genes in 027 ribotype

In addition to the main S-layer SlpA, the 28 other genes containing the cell wall binding domain (CWBD) (Sebaihia *et al.* 2006) were also assessed for intra 027 ribotype variation.

The R20291 sequence (used as a 027 ribotype reference sequence) and 196 (and by assumption 855, BI1 and QCD-32g58) lack a region of genes between *cd2517A* - *cd2520*. Thus, the selected 027 ribotype sequences, and by assumption all 027 ribotype sequences lack the surface protein CD2518, found in this region. Analysis upstream of *cd2518* in 630 reveals an integrase (*cd2519*) suggesting that *C. difficile* 630 may have acquired *cd2518* via an insertion or gene duplication event.

A total of 23 of the 27 remaining CWBD containing gene sequences had >98% identity between *C. difficile* 630 and R20291, four surface proteins did not:

- CD2794 which contains three N-terminal CWBD repeats, a bacterial Ig-like domain (Pfam07523) and a SCP-like extracellular protein domain
    - 95.9% identity.

- SlpA (CD2793) – main S-layer - which contains three C-terminal CWBD repeats
    - 71.7% identity
- Cwp66 (CD2789) - which contains three N-terminal CWBD repeats
    - 86.2% identity.
- CwpV (CD0514) – which contains three N-terminal CWBD repeats and a C-terminal variable tandem repeat region (Emerson *et al.* 2009b)
    - 74.6% identity.

Comparision of the 27 CWBD containing proteins genes between R20291, 196, BI1 and QCD-32g58 genomes reveals 25 are 100% identical, while CD3192 and CwpV are not.

The *cd3192* gene (CDR20291_3048) shares 100% identity between R20291 and 196 and BI1, however the BLAST result of strain QCD-32g58 found only 1182 bp of the full length (1653 bp) CD3192 gene (100% identical up to the truncation). This truncation could result in a foreshortened protein of 393 aa rather than 550 aa, resulting in the loss of one (or two) of the PepSY (Pfam03413) domains. CD3192 contains three N-terminal CWBD repeats and two (or three (using InterProScan)) C-terminal peptidase propeptide (PepSY) domains. PepSY domain(s) show homology to the metalloprotease propeptide region and are thought to regulate proteolytic activity, including those which are surface associated (Yeats *et al.* 2004). No gene was found to match CD3192 in *C. difficile* strain 855.

CwpV (CD0514, CDR20291_0440) appears to be truncated by 79 aa (approximately one R20291 CwpV repeat unit) in *C. difficile* BI1 as a result of a stop introduced at 3274 bp. CwpV in 630 is approximately 495 aa longer than CwpV in R20291, as such a large protein (CwpV) may be a target for the immune system, a reduction in size could therefore provide an selective advantage.

Taken together these data suggest the surface proteins of the 027 ribotype are highly conserved. The conservation of the majority of the CWBD containing proteins

suggests a high degree of selective pressure, and may indicated that these 'minor' CWBD containing species, i.e. not *slpA*, may play a larger role in cell physiology or pathogenesis than previously estimated.

4.3.3 Polyclonal anti- SLP antibody production and analysis (investigational and large scale)

C. difficile surface proteins, particularly the HMW S-layer subunit, have been previously shown to bind to gastrointestinal tissues and to the ECM proteins collagen I, thrombospondin, and vitronectin (Calabi *et al.* 2002). Furthermore, a number of other putative surface proteins have been shown to be involved in adhesion (Hennequin *et al.* 2001; Tasteyre *et al.* 2001a; Hennequin *et al.* 2003).

These data, together with the conservation of the CWBD containing surface proteins, suggest the surface proteins represent a strong vaccine candidate for targeting *C. difficile* colonisation and that adhesion and colonisation is probably multifaceted. This study sought to determine if high titre anti-surface protein immunotherapy could prevent CDI caused by an epidemic ribotype 027 strain in the hamster model.

To develop anti-SLP serum for passive immunisation experiments, crude SLP extracts were first used to raise guinea pig, then after characterisation, ovine polyclonal hyperimmune serum. By using crude low pH surface protein extracts, the multifaceted adhesion of *C. difficile* is partially addressed.

A total of three *C. difficile* strains (two ribotypes) were used for antigen production: *C. difficile* R20291 Stoke Mandeville epidemic strain (Ribotype 027) and VPI10463-high toxin producer (investigational scale) and ARL001 (large scale) strains representing the previous pre-2008 clinically significant 001 ribotype. Prior to immunisation, an assessment was made as to the toxicity of SLPs on Vero cells with either contaminating biologically active *C. difficile* toxin(s) or other harmful protein components which may have a significant effect on the a) population of antibodies generated b) animal wellbeing during immunisation. No toxicity was observed for crude SLP extracts permitting antiserum production to commence.

4.3.3.1 Characterisation of the guinea pig (investigational scale) anti-*C. difficile* SLP immune response.

Using solid-phase SLPs in ELISA, both strains anti-SLP antiserum demonstrated that a strong immune response was obtained (titre >10^5 dilution) (individually R20291 1 x 10^6, VPI10463 5 x 10^5) (Figure 4.3.2).

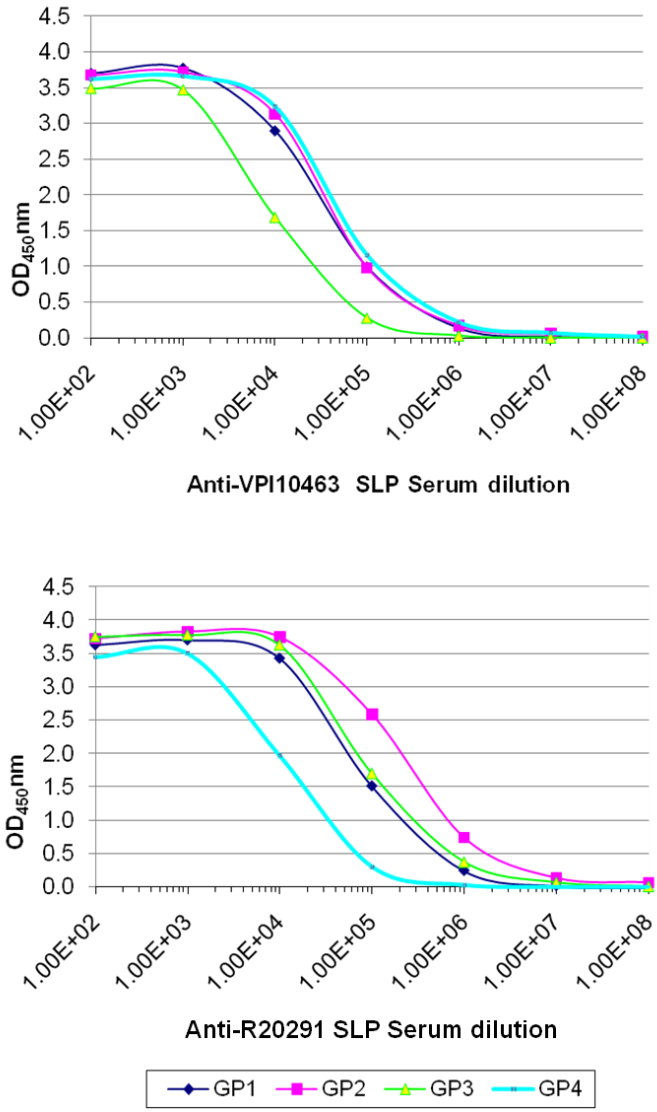


Figure 4.3.2 Guinea pig anti-SLP serum titre ELISA. Four guinea pigs were immunised with 20µg crude low pH SLP extract from *C. difficile* VPI10463 or R20291 strains.

In order to determine which of the specific SLPs the guinea pig antiserum was binding to in ELISA, immunoblots were performed using purified antibodies.

Figure 4.3.3 shows anti-R20291 SLP IgG detected the main S-layer bands of the originating strain, while cross reaction of 630ΔErm and both VPI10463 and ARL001 showed a marked preference for detection of the HMW SLP and a lesser response to the LMW.

Anti-VPI10463 SLP serum detected both the SLP subunits in the originating strain and of ARL001, while cross reacted with R20291 SLPs with similar intensity to the cross reactivity of anti-R20291 with VPI10463 SLPs i.e. main HMW, lesser LMW response. Only the HMW SLP of 630ΔErm cross reacted with anti-VPI10463 SLP IgG, consistent with previous reports and indicating sufficient divergence of SlpA (Takeoka *et al.* 1991; Calabi *et al.* 2001). Both anti-R20291 and anti-VPI10463 SLP IgG detected a 66 kDa band in all SLPs tested and two additional species at approximately 84 and 90 kDa in R20291, VPI and ARL001 SLPs.

The reaction of the other species in immunoblots suggests that other protein species in SLP extracts are capable of generating an immune response. Interestingly, anti-SLP IgG did not react a significant number of protein species <35 kDa, suggesting that these proteins are either not immunogenic or are not presented to the immune system in native SLPs.

slpA of VPI10463, ARL001 and R20291 was sequenced to examine the cross reactivity of guinea pig anti-SLP antibodies. *slpA* of R20291 and VPI10463 share 92.0% identity at base pair level and 89.3% at amino acid level, sufficiently high for immune cross reactivity. VPI10463 and ARL001 are 100% identical at base pair level.

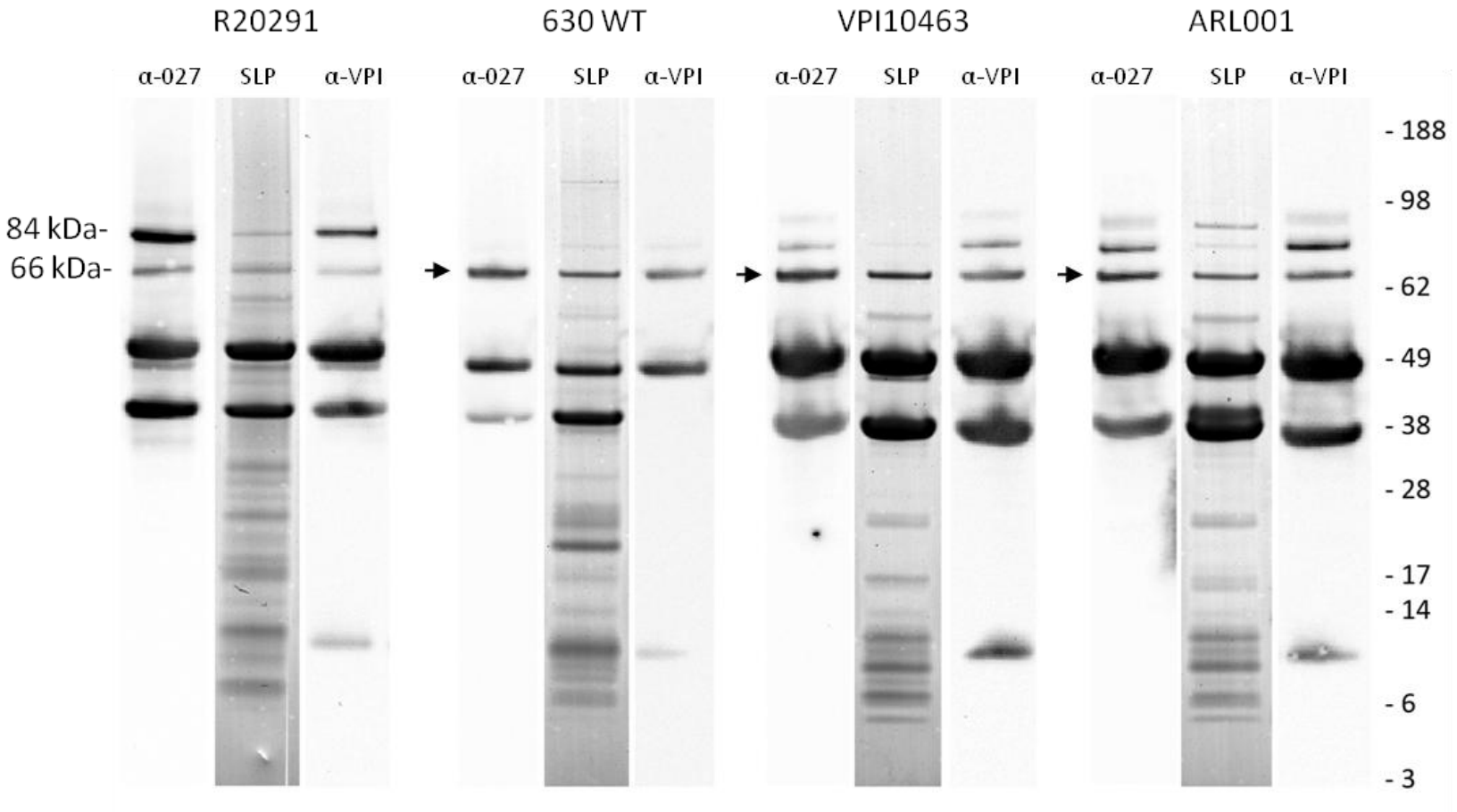


Figure 4.3.3 Immunological cross-reactivity of SLPs from a range of *C. difficile* strains using guinea pig anti-SLP IgG. Low pH SLPs from strains R20291, 630ΔErm, VPI10463 and ARL001 (centre lanes) were blotted with 1 µg/ml guinea pig anti-SLP IgG raised against SLPs from R20291 (left) or VPI10463 (right). Arrows highlight a 66 kDa species detected using all anti-SLP serum.

Binding of anti-SLP antibodies to vegetative cells is key in establishing if anti-SLP antibodies can be used immunotherapeutically. Thus, guinea pig anti-R20291 SLP antibodies were fluorescently labelled, then their binding to vegetative cells examined by fluorescent microscopy. Anti-R20291 SLP antibodies (100 μg/ml) bound to vegetative cells; however, their binding did not decrease upon pre-incubation of cells with 100x concentration of unlabelled antibodies, suggesting that the binding sites for the antibodies are present at a high concentration on the cell surface.

To quantify antibody binding to vegetative cells, whole cell suspensions and SLP extracts (positive control) were assayed by (bacti-) ELISA, whereby bacterial cells were immobilised on solid phase followed by normal ELISA protocols.

Anti-VPI10463 SLP IgG bound to VPI10463 and R20291 SLPs to the same extent, but to 630ΔErm SLPs and 630ΔErm and R20291 vegetative cells approximately 1-log less (Figure 4.3.4A). Anti-VPI10463 SLP IgG appeared to give a poor response to VPI10463 vegetative cells. Anti-R20291 SLP IgG bound to R20291 and VPI10463 SLPs with similar affinity, binding to 630ΔErm SLPs and R20291 vegetative cells was approximately 1-log less, while binding to 630ΔErm cells was a further log less. Anti-R20291 SLP IgG also bound poorly to VPI10463 vegetative cells (Figure 4.3.4B). No binding was obtained when using *C. difficile* toxin B fragment antiserum.

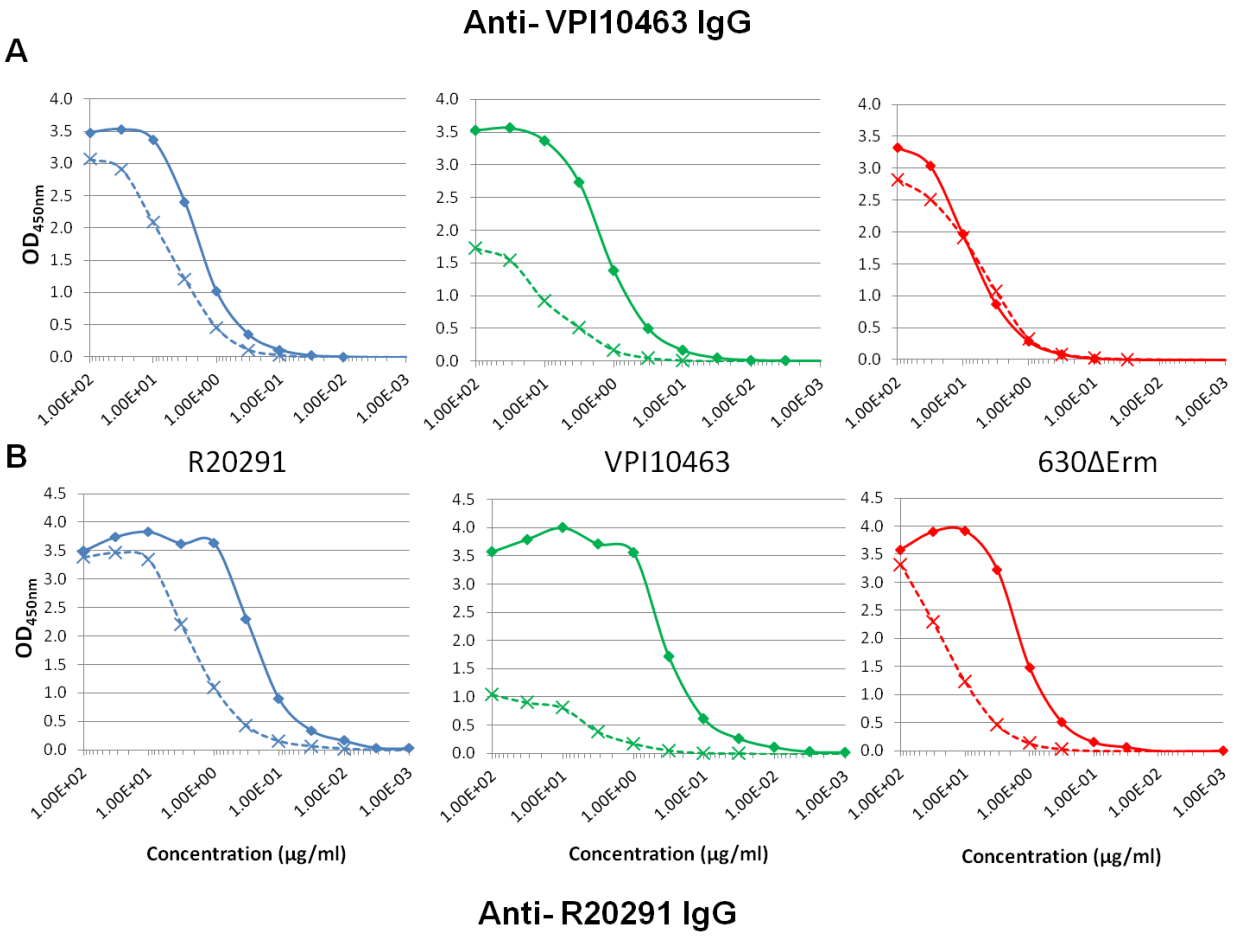


Figure 4.3.4 Cross reactivity of guinea pig anti-SLP IgG to SLP and whole cells of different *C. difficile* strains measured by ELISA. (A) Reaction of anti-R20291 SLP IgG and (B) Reaction of anti-VPI10463 SLP IgG to low pH SLPs (solid lines) and whole cell preparations (dashed lines).

4.3.3.2 Characterisation of ovine (Large scale) anti-SLP antibody response

Taken together, these data suggest that a good high titre antibody response can be generated to crude low pH surface protein extracts of *C. difficile*. The antibodies generated appear to be specific to certain SLPs but bound to vegetative cells. Responses to vegetative cells were good in the corresponding strain but are approximately 1-log less in a differing ribotype, where the predominantly the HMW SLP cross-reacts e.g. anti-VPI10463 SLP IgG on 630ΔErm cells. Due to the amount of anti-SLP IgG required to assess the efficacy of anti-SLP antibodies *in vivo* i.e. passive immunisation studies (10 hamsters given three 1 ml doses of 50 mg/ml anti-SLP IgG (~1.5 g IgG), large scale antiserum production was necessary.

Large scale ovine anti-SLP serum was generated to two *C. difficile* ribotypes, the 027 ribotype (strain R20291) and the clinically prevalent 001 ribotype (Cheknis *et al.* 2009; Bauer *et al.* 2011) (using the Aerobe Reference Laboratory reference 001 ribotype strain, ARL001) rather than the high toxin producing strain VPI10463 (Merrigan *et al.* 2010). The *slpA* gene of *C. difficile* ARL001 strain is 100% identical with VPI10463, thus results from investigational analysis using VPI10463 SLPs should be transferable. However, analysis of the SLPs of both strains reveals subtle differences (Figure 4.3.1) which may decrease transferable results and/or highlight how although *slpA* is identical, the SLPs of VPI10463 and ARL001 differ.

As previously, using solid-phase SLPs in ELISA, both strains anti-SLP antisera demonstrated that a strong immune response was obtained, however only slightly better than guinea pigs (titres: R20291 1 x 10^6, ARL001 1 x 10^6) possibly indicating a limit of immune response had been reached with the crude SLP preparation (Figure 4.3.5).

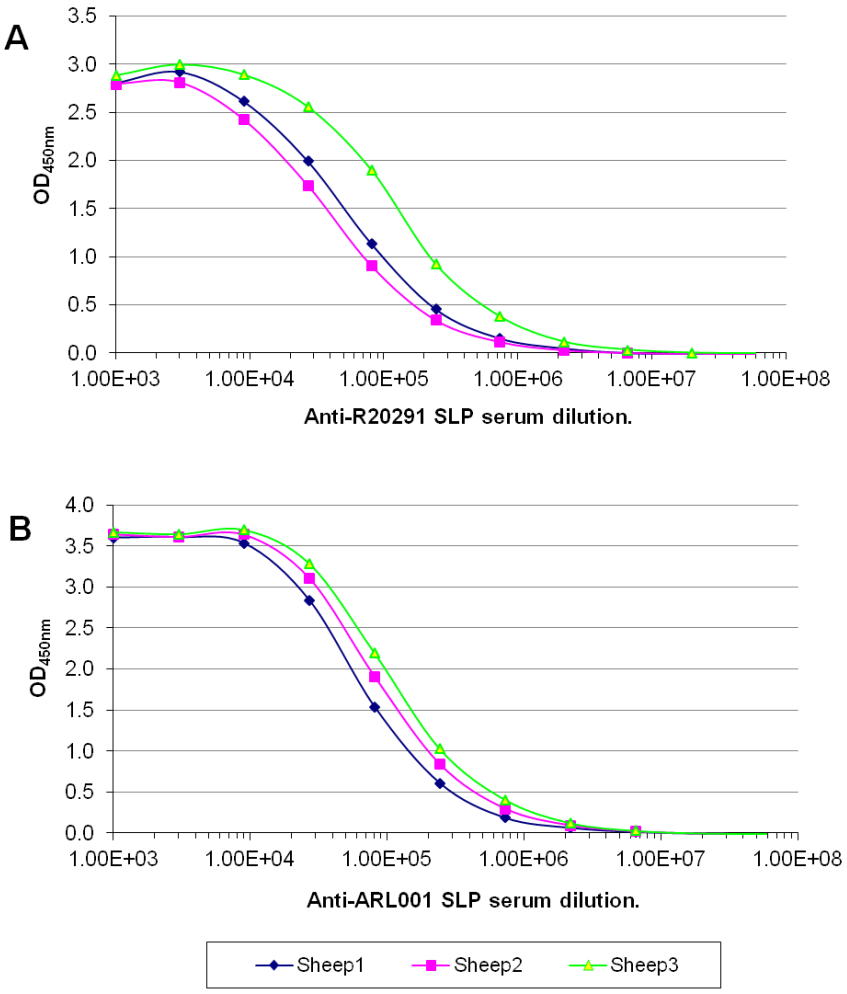


Figure 4.3.5 Ovine anti-SLP serum titre ELISA. Three sheep were immunised with each antigen.

Figure 4.3.6 (α-027) shows that ovine anti-R20291 SLP antiserum detects a pattern of proteins in SLPs similar to guinea pig anti-R20291 SLP IgG. The HMW and LMW SLPs cross reacted between R20291 and VPI10463, however there was a preference for the HMW SLP of ARL001 and, consistent with previous data, only the HMW of *C. difficile* 630ΔErm SLPs was detected. As with guinea pig antibodiesa 66 kDa species gave a good immune reaction in all strains tested including *C. difficile* 630ΔErm. Ovine anti-R20291 also appeared to detect an 84 kDa species in all SLP extracts examined.

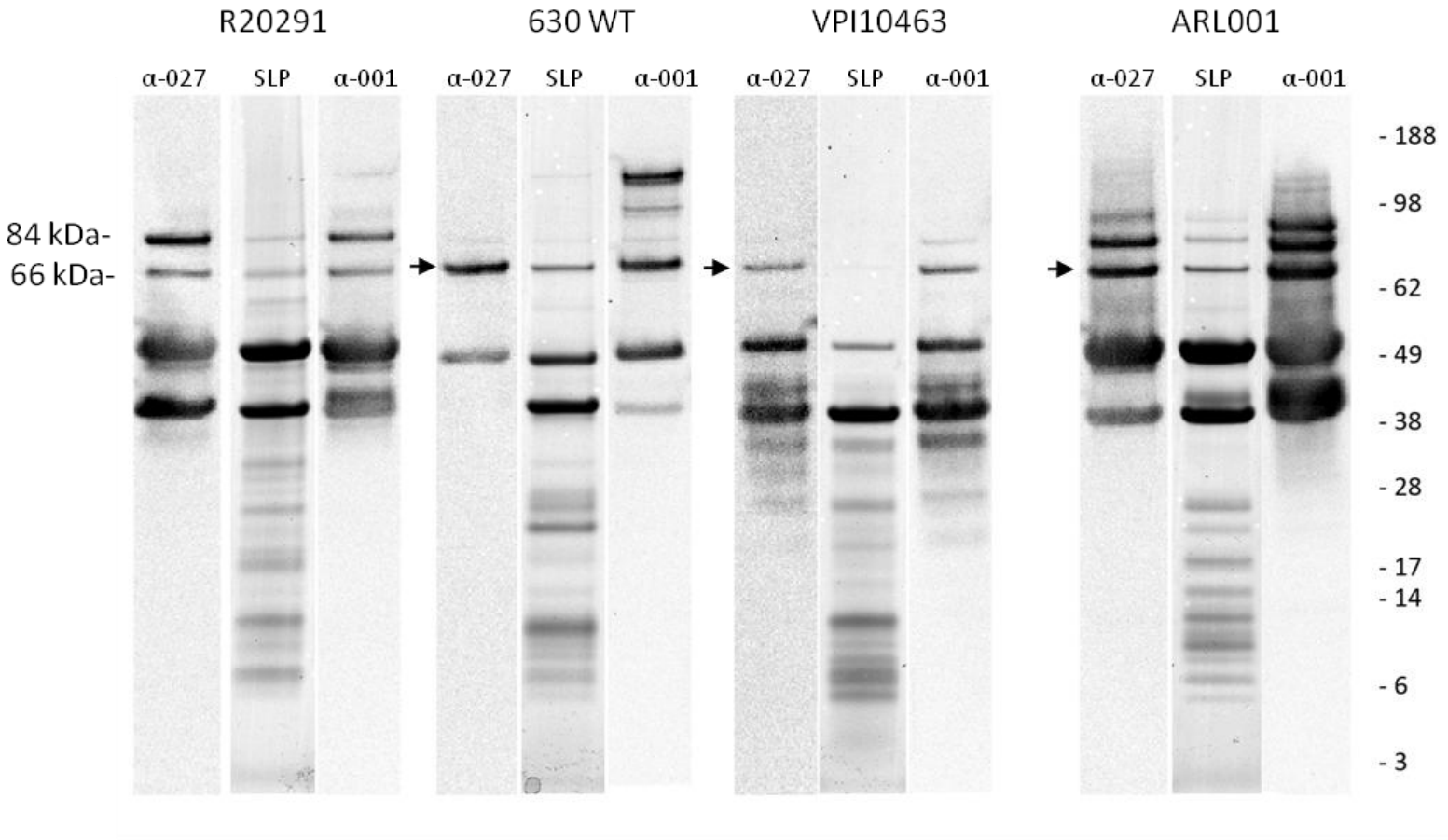


Figure 4.3.6 Immunological cross-reactivity of SLPs from a range of *C. difficile* strains using ovine anti-SLP IgG. Low pH SLPs from strains R20291 (027), 630ΔErm, VPI10463 and ARL001 (centre lanes) were blotted with 1 µg/ml ovine anti-SLP IgG raised against SLPs from R20291 (left) or ARL001 (right). Arrows highlight a 66 kDa species detected using all anti-SLP serum.

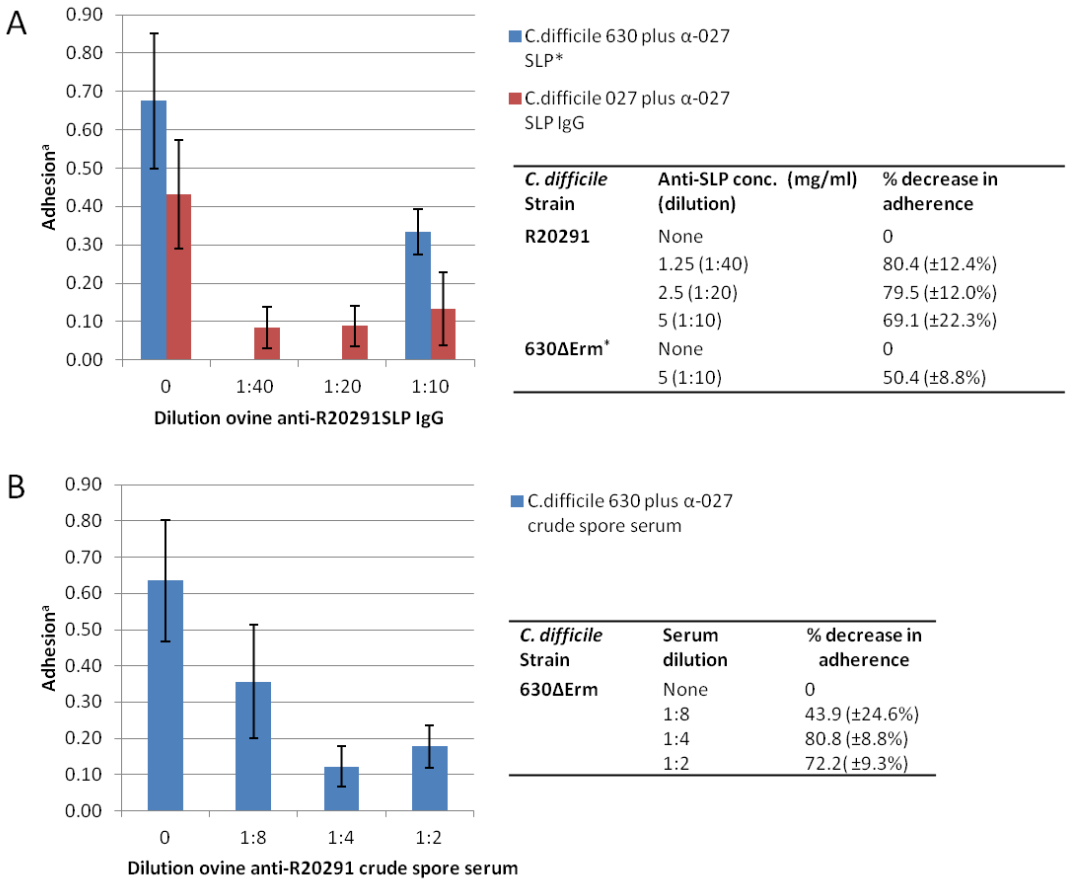
Ovine anti-ARL001 serum detected both the HMW and LMW SLP in all extracts tested, although with a preference for HMW SLP detection of differing ribotype. Similar to ovine anti-R20291 SLP, ovine anti-ARL001 reacted with bands at 66 kDa and 84 kDa. A 115 kDa species was detected with ovine anti-001 in the *C. difficile* 630ΔErm SLP extract, despite this protein not being visible in ARL001 SLP extracts. (Figure 4.3.6, α-001)

The above data supports investigational scale antiserum production analysis, that S-layer extracts can generate a good, high titre, specific immune response. There is also compelling evidence which suggests other 'minor' species also present in SLP extracts are also good at generating an immune response e.g. the 66 kDa species, which are capable of cross ribotype reaction where others e.g. LMW SLP, are not.

4.3.4 Prevention of binding of *C. difficile* cells in a cell adhesion assay

Antibodies raised to low pH SLP extracts of *C. difficile* recognise and bind to vegetative cells *in vitro*. In order to determine if anti-SLP antibodies were able to prevent the adhesion of *C. difficile* to gut epithelial cells, an *in vitro* adhesion assay with Caco2 cells was used.

Ovine anti-R20291 (027) SLP IgG significantly inhibited *C. difficile* R20291 binding at all concentrations assessed (Students T-test, $p<0.001$). However, Anti-R20291 SLP IgG, while reducing adhesion of *C. difficile* 630ΔErm, the difference was not significant compared to control (Students T-test, $p=0.09$) (Figure 4.3.7A) suggesting that anti-SLP antibodies preferentially prevent binding of their cognate strain but are less effective against a differing strain.



C. difficile Strain	Anti-SLP conc. (mg/ml) (dilution)	% decrease in adherence
R20291	None	0
	1.25 (1:40)	80.4 (±12.4%)
	2.5 (1:20)	79.5 (±12.0%)
	5 (1:10)	69.1 (±22.3%)
630ΔErm*	None	0
	5 (1:10)	50.4 (±8.8%)

C. difficile Strain	Serum dilution	% decrease in adherence
630ΔErm	None	0
	1:8	43.9 (±24.6%)
	1:4	80.8 (±8.8%)
	1:2	72.2(±9.3%)

Figure 4.3.7 Inhibition of *C. difficile* binding to Caco-2 cells, by (A) ovine anti-R20291 SLP purified IgG and (B) ovine anti-R20291 crude spore prep serum.

[a] Mean numbers of *C. difficile* organisms adhering to a cell (± standard error).

* Mean from all assays in which anti-R20291 SLP IgG was used against *C. difficile* 630ΔErm (Mean Adhesion[a] = 0.68±0.18, mean adhesion plus anti-R20291 SLP IgG 0.33±0.06, n=7)

Measurements made by random field of view counts of a total of >150 Caco-2 cells vs bacterial cells at x1000 magnification, each adhesion assay was conducted in duplicate.

By comparison, use of an ovine antiserum raised against a *C. difficile* R20291 crude spore preparation, which included spores and vegetative cells, significantly inhibited binding of *C. difficile* R20291 (027), VPI10463 and 630ΔErm to Caco-2 cells (Students T-test $p<0.05$ for R20291, VPI10463 and 630ΔErm) (Table 4.3.2A). Non-specific antiserum (raised to *E. coli* OmpA), used at the same concentration, did not significantly prevent binding of 630 to Caco-2 (Students T-test $p= 0.6411$) (Table 4.3.2B).

Table 4.3.2 Inhibition of various *C. difficile* strains binding to Caco-2 cells by ovine anti-crude spore serum.

A

Strain	Adhesion[a]	Adhesion[a] + anti- spore serum	% decrease in adherence
R20291	0.66 ± 0.07	0.06 ± 0.02	90.9 ±3%
VPI10463	1.03 ± 0.11	0.53 ± 0.08	48.5 ±8%
630ΔErm	0.44 ± 0.06	0.11 ± 0.02	75.0 ±5%

B

Strain	Adhesion[a]	Adhesion[a] + anti-spore serum	Adhesion[a] + anti-OmpA serum
630ΔErm	1.13 ± 0.10	0.29 ± 0.04	1.20 ± 0.11

[a]Mean numbers of *C. difficile* organisms adhering to a cell (± standard error). Measurements made by random field of view counts of a total of >150 Caco-2 cells vs bacterial cells at x1000 magnification, each adhesion assay was conducted in duplicate.

As with anti-SLP antibodies, crude spore serum inhibited binding most when used against the corresponding strain (Table 4.3.2 A) and appeared to significantly inhibit adhesion at a dilution >1:8 (Figure 4.3.7B).

The numbers of bacteria adhering to cells (approximately 0.5 bacteria/cell) was consistent with that found for both the growth medium used (BHI) and cell line tested (Karjalainen *et al.* 1994).

Together, these data suggest that antibodies raised to proteins present in low pH SLP extracts can prevent binding *in vitro*; however antibodies raised to a crude spore-vegetative cell preparation containing other antigen(s) further decreases adhesion of *C. difficile* to caco-2 cells. Furthermore, serum raised to SLPs of one *C.*

difficile strain are effective in inhibiting adhesion in another, although is most effective upon the strain from which the SLPs were derived.

4.3.5 Passive immunisation of hamsters challenged with *C. difficile*

To date, only one study has been performed assessing the ability of anti-SLP antibodies to prevent CDI (caused by a ribotype 001 strain (*C. difficile* R13537) in the hamster model (O'Brien *et al.* 2005). The study found although hamsters succumbed to disease, time-to-death was significantly longer in the antibody group. As a result of this study and data presented in this chapter, the ability of passively administered systemic anti-SLP antibodies to the epidemic strain R20291 (027 ribotype) were assessed for their ability to prevent CDI in the hamster model. In an initial experiment, where hamsters were challenged with 2 x 10^2 cfu/ml *C. difficile* R20291 spores, animals did not succumb to CDI. The experiment was therefore repeated with 2 x 10^5 cfu/ml *C. difficile* R20291 spores, together with an increased time from clindamycin to challenge, as R20291 is clindamycin sensitive.

Figure 4.3.8 shows that hamsters immunised intraperitoneally with three doses (500 mg/kg) of ovine anti-R20291 SLP IgG, succumbed to CDI at a similar rate to un-immunised animals. The data suggests that systemically administered anti- *C. difficile* surface protein extract IgG does not protect against CDI caused by the *C. difficile* R20291 (027 ribotype) strain.

4.3.6 Summary

In summary, low pH surface protein extracts of *C. difficile* are capable of inducing an immune response. This response is specific to certain protein species in SLP extracts but nevertheless recognises whole cells and prevents adhesion to cell lines *in vitro*. The conservation of surface proteins within the 027 ribotype, coupled with the identification of a conserved protein in *C. difficile* SLP extracts recognised by any strain's anti-serum, provides evidence of stable surface based antigens with (immuno)therapeutic potential. However, passively administered high titre antibodies to surface based antigens were not capable of preventing CDI in the hamster model, bringing into question the ability of these antigens to protect against CDI.

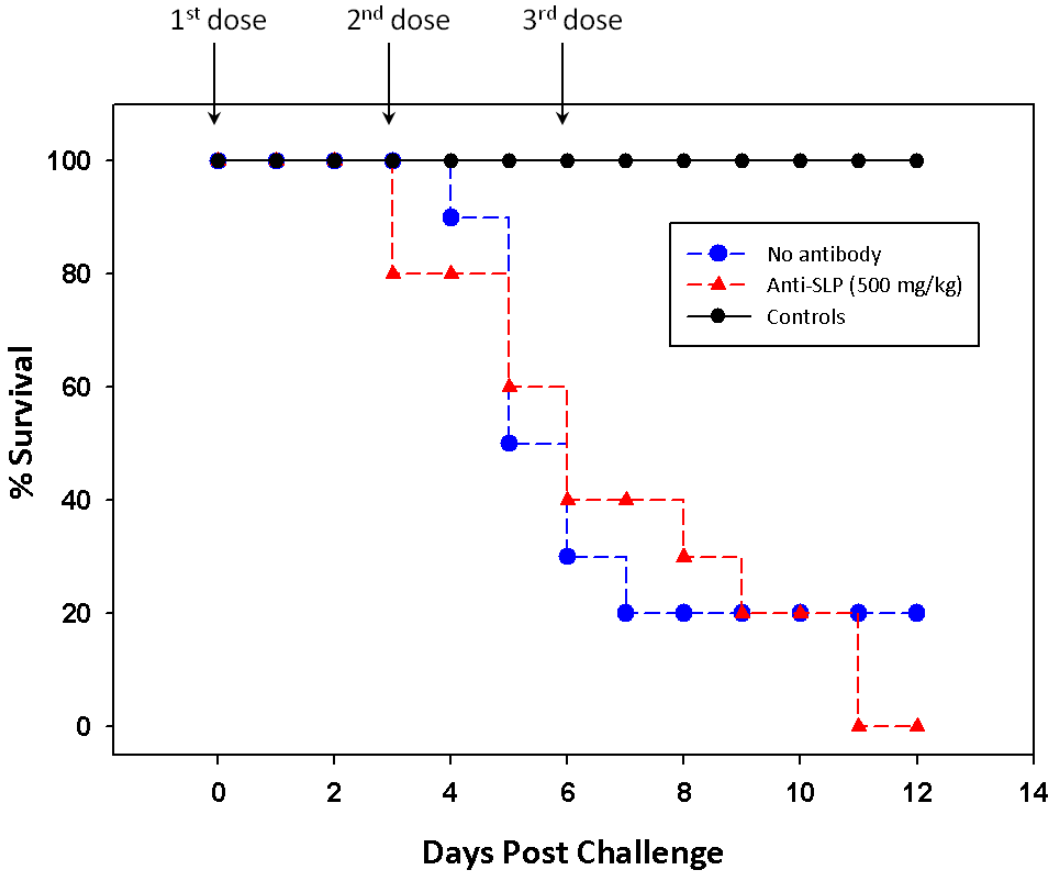


Figure 4.3.8 Ability of passively administered ovine anti-R20291 SLP IgG to prevent CDI in hamsters. 2 test groups of 10 animals were challenged with 2 x 10^5 cfu/ml *C. difficile* R20291 (027 Ribotype). Animals in 1 test group were administered 500 mg/kg of anti-R20291 SLP IgG on day of challenge and 3 and 6 days later. Data show time from challenge to severe disease/death in the hamster model for CDI.

4.4 Discussion

C. difficile infection is primarily a toxin mediated disease. As a result of an inability to mount an effective immune response, passive administration of antibodies appears to be able to neutralise the cytotoxic activity of the toxins (Babcock *et al.* 2006; Demarest *et al.* 2010; Leav *et al.* 2010; Lowy *et al.* 2010)

It appears that certain proteins on the surface of the bacterium can also elicit a systemic immune response in patients with CDI (Pantosti *et al.* 1989; Drudy *et al.* 2004; Wright *et al.* 2008). Moreover, an immune response against selected surface proteins may prevent colonisation (Pechine *et al.* 2007), certainly antibodies against surface proteins can prevent their binding to HT-29 cells (Calabi *et al.* 2002). However, it is currently unclear as to whether an inability to respond to surface proteins predisposes CDI, or whether this highlights the patient's own immunocompromised state e.g. immunosenescence and the need for a therapy to boost such responses.

Once a facility is contaminated with *C. difficile*, infection occurs more readily (Bartlett 2008b). Therefore, there is a putative 'colonisation window' in which the spores germinate and vegetative cells colonise, then produce toxin. *In vitro*, toxin production only occurs once growth is well underway and cells reach stationary phase (Merrigan *et al.* 2010). Therefore, colonisation represents a key milestone in infection. The currently clinically most prevalent 027 ribotype putatively produces more spores which are thought to contribute heavily to its increased virulence (Akerlund *et al.* 2008; Merrigan *et al.* 2010).

To date, there has only been a suggestion that the 'hypervirulent' ribotype 027 has an altered surface possibly increasing the ability to cause infection (Cartman *et al.* 2010). This study sought to understand if there was any evidence of surface based differences between *C. difficile* strains e.g. the 027 ribotype and to assess anti-surface protein immunotherapy against CDI caused by a 027 ribotype strain (R20291). This study suggests that a large majority of the surface proteins in 027

ribotype strains appear the same as non-epidemic strains. Furthermore, antibodies raised to the surface proteins of one strain can inhibit the adhesion of another, but are not protective against the cognate strain in the hamster model.

4.4.1 Surface proteins of PCR ribotype 027 isolates

4.4.1.1 Surface layer protein extraction and sequencing of *slpA*

There appears to be some inconsistency with the molecular weights of the main SLPs observed in this chapter and those reported previously. The 001 ribotype S-layer bands are expected at 34 kDa and 50 kDa (but found at 39 kDa and 49 kDa in this chapter) while 630 (ribotype 012) S-layer bands are expected at 33 kDa and 47 kDa (but found at 40 and 47 kDa in this chapter) (Calabi & Fairweather 2002; Eidhin *et al.* 2006). Recently, SDS-PAGE analysis of S-layer bands of 027 ribotype strains were found to match their predicted MW (34 kDa and 44 kDa) (Spigaglia *et al.* 2011a). However, SLP extracts from *C. difficile* R20291 in this chapter found the LMW and HMW SLP bands at 40 kDa and 50 kDa respectively. Reasons for the migration differences are unknown, but likely to be due to differences in the SDS-PAGE system, e.g. gel type and concentrations used.

Eidhin *et al.* (2006) suggested that the sequence of *slpA* from different isolates of the same ribotype are all identical. The sequencing of *slpA* from the regional 027 isolates in this chapter and work by Spigaglia *et al.* (2010; 2011a) confirms this. Both confirm how *slpA* sequencing is insufficient in discriminating one 027 isolate from another but appears to compliment PCR ribotyping. However, the 027 ribotype although genetically highly uniform, can be divided into a number of different subtypes based on different typing methods e.g. DNA microarray or multilocus variable-number tandem-repeat analysis (MLVA) (Stabler *et al.* 2006; Killgore *et al.* 2008; Stabler *et al.* 2009; Tanner *et al.* 2010). However, the applicability of these techniques, in a clinical outbreak setting for example, may be limited.

The 027 ribotype is reportedly a robust toxin producer (Merrigan *et al.* 2010), if therefore increased toxin production had a surface based determinant, one may

observe differential expression of surface proteins between a high and low toxin producer. In this chapter, the SLPs of 630ΔErm (low toxin producer) appeared little different to R20291 or VPI10463 (high toxin producer). The earlier toxin release of selected surface proteins knockout does, however, highlight how the bacterial surface and toxin production may be linked, especially *in vivo*.

Therefore, to date, no specific cell surface based reason for hypervirulence has been suggested. The finding that 001 ribotype and 027 ribotype *slpA* sequences are similar (genetically and immunogenically (Spigaglia *et al.* 2011a; Spigaglia *et al.* 2011b) and this work) suggests that these common CDI causing strains may have a unique surface particularly well adapted to colonisation. The genes encoding *slpA* and *cwp84* are upregulated in response to various stresses (Deneve *et al.* 2008; Emerson *et al.* 2008; Deneve *et al.* 2009a) which may suggest a 'shoring up' of the cell surface in defence. Thus, there may be differences in gene expression, e.g. up/down regulation or recombination events with other *slpA* homologs, which occur *in vivo,* and result in an increase in surface based adhesion or virulence factor expression, particularly in hypervirulent strains. For example, the surface layer proteins (encoded on a 54 kbp chromosomal DNA locus "sap island") of *Campylobacter fetus* undergo rearrangement to up-regulate virulence, for antigenic variation and to enter a latent state (Tu *et al.* 2005).

4.4.1.2 Other species present in surface protein extracts

While the HMW and LMW SLPs undoubtedly play key roles in cell physiology and pathogenesis, the plethora of 'minor' hitherto uncharacterised proteins found on the cell surface (Wright *et al.* 2005) are likely to play key roles in pathogenesis such as cell signalling, matrix degradation and adherence.

A 66 kDa protein, identified as CD2791 (aka Cwp2), is found in every surface protein extract to date, except in strain 167 (Calabi & Fairweather 2002). This protein is highly conserved and reacts with antibodies raised to any strains SLPs. CD2791, found during normal growth and discussed further in Chapter 5, is putatively part of a polycistronic transcript including the putative 66 kDa heat shock adhesin Cwp66

and an unknown ORF (CD2790) (Savariau-Lacomme *et al.* 2003). CD2791's role in pathogenesis or cell physiology therefore warrants further investigation.

An 84 kDa protein in SLP extracts, particularly in R20291, could be immature SlpA similar to that identified in Cwp84 KO SLPs (Kirby *et al.* 2009). The greater prominence of this band in R20291 could suggest *slpA* upregulation in R20291, a known response to bacterial stress (Deneve *et al.* 2008; Emerson *et al.* 2008; Deneve *et al.* 2009a).

The extract of *C. difficile* 630ΔErm has a prominent 120 kDa band, identified as CwpV, not visible in the extracts of R20291 or VPI10463. The *cwpV* gene is under phase variable expression (Emerson *et al.* 2009b), thus a decrease in its expression, potentially in R20291, could indicate how it may be detrimental to have large amounts of CwpV presented on the cell surface.

In summary, this study found no surface based reason for increased virulence in the 027 ribotype using low pH SLP extraction. This may be due to no differences being present in hypervirulent strains, the severely limited methods used in this study or the necessary surface based determinants not being expressed *in vitro* or under the conditions used.

4.4.1.3 In silico analysis of other cell wall binding domain containing surface protein genes in 027 ribotype

Analysis of *C. difficile* 027 genomes has mainly centred on comparison of antibiotic resistance, genes related to survival and unsurprisingly toxin related genes (Stabler *et al.* 2009). However, little has been discussed regarding differences in the 28 paralogs of *slpA* which contain a CWBD found in the genome of *C. difficile* 630 (Sebaihia *et al.* 2006), amongst 027 isolates.

Comparison of R20291 and 630 revealed that only three genes were found to be significantly different: *slpA*, *cwp66* and *cwpV*. It is well documented that *slpA* is variable given its prominence on the cell surface and the need to overcome the

host immune response (Karjalainen *et al.* 2001; Calabi & Fairweather 2002; Eidhin *et al.* 2006). Reasons behind the variability of *cwp66* is less well understood, but is thought to be as a result of gene clustering with *slpA* whereby both undergo recombinational events and/or strong environmental selective pressure (Lemee *et al.* 2005). Results by Waligora *et al.* (2001), and particularly results in Chapter 5, are in support of *cwp66* being under selective pressure because Cwp66 is an adhesin. CwpV, as aforementioned, is a large phase variable multi-tandem repeat containing protein. Its function and reason for variability are unknown but it may be involved in detachment from the host or bacteria-bacteria interactions (Emerson *et al.* 2009b). The differing size of *cwpV* amongst *C. difficile* strains may suggest it also is under selective pressure.

A key question is perhaps, why are there so many CWBD containing proteins with no apparent function in the *C. difficile* genome? The S-layer of *Campylobacter fetus* has one promoter but a number of S-layer genes which undergo genetic rearrangement (Thompson 2002; Grogono-Thomas *et al.* 2003). It is interesting to speculate that *slpA* of *C. difficile* has undergone an evolutional progression similar to the *C. fetus sapA* gene. In that initially one S-layer gene may have been present, however successive tandem duplications and positive selection for SLP antigenic variation resulted in a range of S-layer like genes, a pool of SLP genes from which rearrangement could occur.

In summary, the most prominent result from 027 *slpA* paralog analyses is the 100% similarity of the majority of the genes in R20291, BI1, QCD34g-58 and 885. The conservation of the CWBD genes of *C. difficile* ribotype 027 strains suggests genetically every 027 has the same cell surface, although this hypothesis requires further investigation. However, if all strains during an epidemic have the same cell surface antigen(s) the prognosis for treatment via surface based, e.g. anti-colonisation treatment is good.

4.4.2 Immunotherapy of *C. difficile* infection with polyclonal antibodies raised to cell surface proteins.

Antibodies directed to *C. difficile* toxins are currently under development and appear to be able to prevent disease both in the hamster model (Kink & Williams 1998; Babcock *et al.* 2006) and in limited clinical cases or clinical trials (Wilcox 2004; Koulaouzidis *et al.* 2008; Leav *et al.* 2010; Lowy *et al.* 2010), although much work is needed (Abougergi *et al.* 2010).

Antibodies to the surface proteins of *C. difficile* are raised during CDI (Drudy *et al.* 2004; Pechine *et al.* 2005b; Wright *et al.* 2008). However, there are limited studies regarding the protective nature of this immune response. Passive administration of anti-SLP antibodies prolonged the life of hamsters, while active immunisation using SLPs failed to induce a protective response (O'Brien *et al.* 2005; Ni Eidhin *et al.* 2008). In mice, select surface proteins were successful in decreasing colonisation, but the protective effects, in terms of CDI development, of this decrease were not assessed (Pechine *et al.* 2007). Therefore, the ability of passively administered anti-SLP antibodies to protect against CDI, in the hamster model, caused by the epidemic 027 ribotype was investigated.

4.4.2.1 Characterisation of investigational scale ant-SLP antiserum

Small mammals have been used to generate antiserum from both crude surface protein extracts and selected SLPs in a number of studies (Cerquetti *et al.* 1992; Calabi *et al.* 2001; O'Brien *et al.* 2005; Janoir *et al.* 2007). In this study, small mammals were used to generate investigational scale amounts of antiserum which enabled comprehensive analysis of titres, responses and toxicity of the crude SLP antigens.

It appears that SLPs are not toxic to Vero cells, although elements of SLPs bind to and cleave extracellular matrix (ECM) proteins (Calabi *et al.* 2002; Janoir *et al.* 2007). Therefore, data in this chapter supports Sanchez-Hurtado & Poxton (2008), who also found low pH SLPs to be non-toxic to Vero and Caco2 cells.

Due to differences in immunisation protocols and antiserum titre estimation method variation, comparison of titres obtained in this chapter with previous studies is difficult e.g. whole *C. difficile* cell agglutination (O'Brien *et al.* 2005) or Log OD_{405} (Ni Eidhin *et al.* 2008). Although, Ni Eidhin *et al.* (2008) also found that differences in immunisation protocols e.g. SLP dose, appear to do little to affect ultimate antibody titres, suggesting that once a maximal response is reached it cannot be increased. However, data in this chapter suggest that a high titre hyperimmune response to *C. difficile* SLPs (with an adjuvant) is possible and recognises vegetative cells.

Purification of SLPs via HPLC appears to remove their ability to self-assemble (Takumi *et al.* 1991), thus an unknown co-factor (putatively present in urea based extracts) is suggested to be required for S-layer array assembly. This would be consistent with AFM data (Section 3.3.9.3) which suggested low pH SLPs do not self-assemble on a Mica surface. It is unknown as to whether this ability to form arrays on surfaces *in vitro* affects the ability to generate (protective) immune responses. It is likely then, that it is the structure of the SLPs in solution (*in vitro*) which controls to what extent, to which specific SLP(s) an immune response is generated and whether the obtained immune response is protective or not.

4.4.2.1.1 Cross reactivity of SLPs in investigational scale serum immunoblot

Previous reports have suggested that only the HMW SLP cross reacts when blotted with antiserum generated from a differing strain (Takeoka *et al.* 1991; Calabi *et al.* 2001). However, blots in this chapter revealed weak detection of the LMW SLP from a differing ribotype. Sequence analysis of *slpA* between R20291 and VPI10463 revealed 92.0% identity, explaining the cross reactivity in western blot in this study. Cross reactivity between both HMW and LMW SLPs has been suggested by McCoubrey & Poxton (2001), who suggest cross reactivity occurs within the same 'serotype'. However, ribotype 001 (VPI10463 & ARL001) belongs to serogroup G whereas 027 is unclassified (Karjalainen *et al.* 2002). The high sequence identity (89.0% identity rather than 92.0% identity as found in this chapter) and immunological conservancy between 027 and 001 ribotypes has also been reported

by Spigaglia *et al.* (2010; 2011b), who suggest that surface based virulence determinants of these ribotypes may be linked.

The 66 kDa CD2791 protein cross reacted with all strains antiserum, consistent with its high conservation at gene level (98.1% identity at DNA level between 630 and R20291). Analysis of patient serum found the majority of patients raised antibodies to CD2791 (aka Cwp2) (Wright *et al.* 2008). Moreover, ELISA data in Chapter 4 Section 5.3.3.6 suggests that a significant proportion of the immune response to SLPs is to CD2791. CD2791 appears to be highly immunogenic and able to cross all strain barriers. The presence of a (hyperimmune) response to other 'minor' species in SLP extracts confirms that species other than the HMW and LMW SLP are capable of generating an immune response.

The majority of the proteins with a molecular weight less than the LMW SLP in low pH SLP extracts appear to be fragments of the S-layer subunits (Wright *et al.* 2005), However, despite their prominence in SDS-PAGE, antiserum to SLPs raised in this chapter found that no immune response was generated to them, but these proteins appear to be immunoreactive against CDI patient sera (Wright *et al.* 2008). The later result would suggest that these fragments only become immunogenic as a result of CDI, rather than through the generation of hyperimmune serum with SLP extracts.

4.4.2.1.2 Binding of anti-SLP to vegetative *C. difficile* cells.

In order to quantify the binding of guinea pig anti-SLP antibodies to whole cells, rather than agglutination, an ELISA with immobilised cells was developed based on methods used for *Streptococcus sanguis* and *Moraxella bovis* (Elder *et al.* 1982; Prieto *et al.* 2003). The anti-SLP serum reacted approximately 1-log less with whole cells than to corresponding SLPs. The reduced binding to vegetative cells compared to SLPs may have several explanations:

- Spatial organisation of the SLPs. When using the corresponding strain, the LMW SLP could react. When using a different strain, only the underlying HMW could react, together with other cross reacting proteins e.g. CD2791.
- The antiserum was raised to low pH extracted SLPs, which may not have the same structure as SLPs presented on the surface of the bacterium.
- The type of bacterial culture used may have differentially expressed reacting SLPs.
- Antibodies do not adhere to *C. difficile* cells as well as to protein (SLPs) coating the ELISA plate wells.
- A lack of cells adhering to solid phase.
- SLPs sloughing off the bacteria and sticking to plate giving the false impression of whole cell binding when only SLPs are present.

Given the ability of anti-SLP antibodies to bind whole cells of different strains theoretically, with enough antibodies any strain anti-SLP antiserum could be used immunotherapeutically against any other strain.

4.4.2.2 Characterisation of large scale anti-SLP antiserum

Ovine polyclonal (polyvalent) antibodies are proven to be effective in the neutralisation of snake venom and have been widely used (Lavonas *et al.* 2009). Furthermore, sheep antibodies are well tolerated and have good stability in humans when administered passively (Landon *et al.* 1995). Ovine antiserum therefore represents a good immunotherapeutic option for the prevention of CDI.

The ability of sheep to develop antibodies to low pH extracted SLPs appeared to be very similar to that of guinea pigs, both in terms of titre and specific protein species detected in immunoblot. This suggests that the ability of the SLPs to generate an immune response is similar in these two mammalian species.

4.4.2.3 Prevention of binding of *C. difficile* to Caco-2 cell lines.

Caco-2 cells have been used previously to determine both putative adhesion related proteins (Eveillard *et al.* 1993; Karjalainen *et al.* 1994; Cerquetti *et al.* 2002) and to identify specific adhesins in *C. difficile* e.g. Cwp66 (Waligora *et al.* 2001) and

GroEL (Hennequin *et al.* 2001). This study sought to identify if antibodies to surface protein extracts could prevent binding *in vitro*, thereby providing support for *in vivo* passive immunisation assessment. Results in this study suggest that anti-SLP antibodies prevent adhesion to caco-2 cells, whereby inhibition of *C. difficile* adhesion is best achieved with a corresponding strain antiserum. When a differing strain is used, binding inhibition is an order of magnitude less. Adhesion appears to be prevented most using antibodies raised to a whole cell and spore preparation rather than to just SLPs, indicating that either a) other factors present on the cell surface, e.g. carbohydrates, are involved in adhesion, or that b) the presentation of proteins (surface or otherwise) on the bacterium is different, resulting in a increased ability to prevent adhesion. In humans, IgM responses to EDTA surface extracts (which include carbohydrate elements) were more than double responses to low pH SLPs, suggesting an factor on the *C. difficile* surface , removable with EDTA, affects immune response (Sanchez-Hurtado *et al.* 2008).

Any surface protein based immunotherapy should therefore ensure it is representative of surface based antigens expressed during CDI *in vivo*. Hyperimmune serum generated, particularly in this chapter, is thus heavily dependent on the state of the cells *in vitro* which may or may not be representative of the cells state *in vivo* during CDI.

4.4.2.4 Passive immunisation of hamsters with anti-SLP

As *C. difficile* surface proteins can induce an immune response and a mucosal immune response to toxins does not appear to be necessary for protection, passively administered (systemic) anti-SLP antibodies were assessed for their ability to protect hamsters against CDI caused by the 'hypervirulent' *C. difficile* R20291 (027 ribotype) strain. Immunised hamsters developed CDI at the same rate as non immunised animals, suggesting passively administered systemic anti-SLP antibodies are non protective; explanations for the lack of protection are multi-factorial.

Reports investigating the human serum antibody response to *C. difficile* surface based antigens struggle to find a significant difference between cases, carriers or

controls (Drudy *et al.* 2004; Pechine *et al.* 2005a; Sanchez-Hurtado *et al.* 2008), suggesting that a serum immune response (or lack of) to surface proteins is not connected with CDI progression i.e. CDI occurs regardless of a serum anti-surface protein immune response. Circulating IgG (to surface proteins) may therefore not be protective due to a lack of access to *C. difficile* vegetative cells prior to toxin mediated damage. Anti- *C. difficile* colonisation antibodies must therefore reach, bind and prevent colonisation before the toxins are produced. A proposed mechanism of antibody mediated toxin neutralisation is the leakage hypothesis i.e. toxins damage the gut such that circulating antibodies are released, or that the toxins cause damage to the mucosa, such that the toxins themselves then become exposed to circulating antibodies.

Given that CDI is largely *C. difficile* toxin mediated, if gut damage has occurred (to allow circulating antibodies access to the gut) then anti-colonisation antibodies may be too late. However, an intestinal transport system has been proposed and may mediate translocation of IgG to the intestinal lumen (Yoshida *et al.* 2004) which would allow binding to *C. difficile* cells in the absence of epithelial damage. Furthermore, although secretory IgA (sIgA) is significantly less sensitive to proteolytic degradation by intestinal enzymes than IgG, some bacterial species produce anti-immunoglobulin proteases enzymes (Holmgren *et al.* 1992; Parsons *et al.* 2004). Thus, in this particular case, anti-SLP IgG may be degraded before inducing any protective effect. In summary, circulating IgG to *C. difficile* surface proteins may not reach the required location, at the correct time and in sufficient amounts to be able to prevent colonisation.

Therefore, an approach to induce/increase the mucosal sIgA response may be more applicable to *C. difficile* surface based antigens and anti-colonisation immunotherapy. The importance of mucosal responses, (particularly) secretory IgA, has been proven in other gastrointestinal pathogens such as Cholera and *E. coli* (Corthesy & Spertini 1999; Neutra & Kozlowski 2006). Pechine *et al.* (2007), using rectal immunisation, induced a (putatively mucosal) response to a cell surface extract antigen which decreased colonisation. However, the ability of this

colonisation decrease to prevent CDI or toxin mediated damage was not discussed. This raises two key questions: To what percentage must colonisation be decreased before CDI is resolved (or prevented)? Can immunotherapy ever reach those levels?

As discussed by Ni Eidhin *et al.* (2008), the non-protective nature of anti-SLP responses may be due to the severity of CDI in the hamster model as opposed to the mouse model. Unfortunately, no investigations were made into whether the anti-SLP IgG administered in this chapter, exerted an, albeit non-protective, anti- *C. difficile* colonisation effect. The sensitivity of the model may therefore not reflect the treatment efficacy, perhaps in mice (Chen *et al.* 2008) or in humans a lower bacterial load may be sufficient to resolve CDI.

The differential presentation of antigens, is likely to affect to which protein(s) a response is generated and to what extent that response is protective. There are therefore two assumptions in this respect:

Firstly, that low pH SLPs used for hyperimmune serum generation are representative of those presented *in vivo*, during CDI, and to which an immune response offers protection. However, Torres *et al.* (1995) found rectal immunisation with a whole cell preparation did not provide hamsters with protection, suggesting that even in the native confirmation, surface protein responses are likely to be non-protective or that anti-surface protein systemic immune response(s) do not reach the gut in time, in high enough concentration. It is also possible that during CDI, different proteins play different roles but may still generate immune responses, for example, the type I pilus system of *Aeromonas salmonicida* contributes to host colonisation but not invasion (Dacanay *et al.* 2010). *C. difficile* may therefore divert the immune system by having highly reactive surface protein(s) which are not involved in CDI or establishing colonisation.

Secondly, the assumption that the mechanism for measuring the immune response e.g. ELISA, is predictive of protection. Antibody ELISA titre and antibody quality i.e. the ability to prevent CDI, is not necessarily linked as toxin neutralising titres

correlate more closely with protection than ELISA titre (Torres *et al.* 1995; Giannasca *et al.* 1999). Therefore, anti-SLP antibody responses in this chapter appear to be 'high-titre' (as measured by ELISA), however there may be insufficient protective antibodies in the antibody preparation.

Finally, the inability of anti-SLP IgG to protect against CDI in the hamster model may be due to the increased virulence of the *C. difficile* R20291 challenge strain used. Although an initial experiment failed to cause disease (suggesting decreased virulence), the pathology of CDI in hamsters is reported to be more severe with shorter time to death (48 hr) compared to the *C. difficile* 630 strain (102 hr) (Razaq *et al.* 2007; Goulding *et al.* 2009).

4.4.3 Immunotherapeutic approaches of CDI using surface proteins as antigens- perspectives

The treatment of *C. difficile* has sparked a number of different therapeutic regimes including toxin binding resins and a range of new antibiotics (Bartlett 2009). However, clinical trial failure, high costs and concerns for resistance may limit the use of these new treatments, hence immunisation (both passive and active) against *C. difficile* disease could represent viable therapeutic and preventative strategies for *C. difficile* infection (Koo *et al.* 2010). To date, the most successful immunisation (active or passive) strategies appears to be derived from toxoided forms of *C. difficile* toxins (Fernie *et al.* 1983; Kotloff *et al.* 2001; Sougioultzis *et al.* 2005).

Data in support of proteins, other than the toxins, eliciting an immune response (Pantosti *et al.* 1989; Drudy *et al.* 2004; Wright *et al.* 2008) creates an opportunity to understand whether immunotherapeutic or vaccination strategies using surface proteins (individually or together) are feasible. For example, co-administration of four surface based antigens of *S. aureus* afforded high levels of protection against invasive disease or lethal challenge with human clinical *S. aureus* isolates in mice, yet only afforded partial protection when the same antigens were administered individually (Stranger-Jones *et al.* 2006).

The variability of patient populations has made determination of whether an immune response to surface proteins is protective, difficult, particularly given possible pre-exposure to *C. difficile* (Viscidi *et al.* 1983; Drudy *et al.* 2004; Sanchez-Hurtado *et al.* 2008). Pantosti *et al.* (1989) suggested that a prolonged presence of *C. difficile* in the gut is required to elicit a strong antibody response to cell surface antigens, thus these antibodies are probably non-protective. It is also likely that immune sampling i.e. immuno-tolerance (Holmgren & Czerkinsky 2005) may prevent excessive reactions to *C. difficile* (surface proteins) as a transient inhibitor of the gut.

Antibodies are protective against antigens derived from colonisation factor antigen I (CFA/I), an *E. coli* fimbriae subunit (Rudin *et al.* 1996), or against Bordetella colonisation factor A (BcfA) an outer membrane immunogenic protein (Sukumar *et al.* 2009), which supports the hypothesis that anti-surface protein vaccines can prevent disease. Moreover, a recent study by Jarchum *et al.* (2011) found that administration of a TLR5 agonist (Salmonella-derived flagellin), prevents CDI in the mouse model, supporting anti- *C. difficile* surface protein approaches.

Passively administered antibodies to S-layers have been proven to partially protect mice against lethal challenge with *Campylobacter fetus* (Blaser & Pei 1993). Moreover, mice immunised with purified S-layer from *Bacteroides forsythus* then challenged with *B. forsythus* cells did not develop any abscesses (Sabet *et al.* 2003), which suggests that the immunisation with S-layer proteins can confer protection against subsequent infection. Although, both of these organisms do not produce toxins which replicate many aspects of the disease, as *C. difficile* toxins do (Borriello *et al.* 1988).

The key to (passive) anti-colonisation immunisation may be in the treatment of relapse/recurrent patients, as patients post-CDI, pre-relapse, are likely to have un-germinated spores in the GI tract. *In vitro* immunoglobulin (IVIG) against toxins has gained recognition in the treatment against relapse/recurrent CDI (Babcock *et al.* 2006; Taylor *et al.* 2008; Leav *et al.* 2010). Anti-colonisation therapy would

therefore prevent germinated vegetative cell colonisation 'resolving' CDI before recurrence. However, anti-Spore preparations rather than anti vegetative cell preparations may be better suited to prevent relapses in a patient continually inhabited by un-germinated spores. Immunisation of rabbits with non-toxigenic *Bacillus anthracis* spores provides protection against toxigenic strains, moreover, passively administered anti-spore IgG protected mice from intraperitoneal challenge with a lethal dose of fully virulent *B. anthracis* (Enkhtuya *et al.* 2006).

4.4.4 Conclusion

The vast similarities of the surface proteins within the 027 ribotype suggest that, should a protective immune response be generated to one strain within the ribotype, protection should be provided against another. If anti- *C. difficile* colonisation immunotherapy is to be successful, a highly protective rather than just high-titre ELISA response must be obtained. Careful selection of proteins key to adhesion and colonisation, but not necessarily highly immunogenic, will be key in this respect, together with methods of antibody administration such that (passively administered) antibodies reach the gut at the optimal point in infection and in sufficient quantity. Furthermore, understanding how or more specifically when, toxins are produced *in vivo* during infection, will be key to establishing if anti-colonisation therapy can prevent CDI. Given that inhabitation of *C. difficile* in the gastrointestinal tract does not appear detrimental; certainly when a normal microbiota is present, prevention of colonisation may therefore be unnecessary in a toxin mediated disease. The commercial application of immunisation against surface proteins is probably best suited to the prevention of relapses or used in conjunction with anti-toxin therapies.

CHAPTER 5 –
Selected *C. difficile* surface protein knockout

5 Chapter 5 - Selected *C. difficile* surface protein knockout

5.1 Introduction

5.1.1 Genetic manipulation of *C. difficile*

Effective genetic manipulation tools are required to understand the molecular basis of pathogenicity and utilise the amassed knowledge obtained from *C. difficile* genome sequencing. However, genetic modification of *C. difficile* and Clostridia is hampered by difficulties in conducting conjugation and transformation experiments as a result of an efficient restriction and modification (RM) system. The RM system comprises of a methyltransferase and restriction endonuclease pair. Restriction endonucleases cleave foreign, e.g. viral DNA, while methyltransferases protect host DNA (by methylating specific bases, which protrude into the major groove of DNA at the restriction enzyme binding site and prevents cleavage)(Wilson & Murray 1991).

Despite the limitations, a number of genetic manipulation reports exist in *C. difficile* including an occurrence of electroporation of the PaLoc from a toxigenic to a non-toxigenic strain (Ackermann *et al.* 2001). Progress in *C. difficile* genetic manipulation occurred through development of plasmids able to transfer, usually by conjugation, into *C. difficile* (*E. coli–C. difficile* shuttle vectors) (Purdy *et al.* 2002).

The knockout of the glycerol dehydrogenase (GldA) by Liyanage *et al.* (2001) utilised PCR fragments of the *gldA* gene, cloned into a conjugative vector (pMTL31 (Williams *et al.* 1990)) which interrupted *gldA* by homologous recombination. The inactivation of two putative response regulator genes, *rgaR* and *rgbR* (O'Connor *et al.* 2006) and a *codY* mutant (Dineen *et al.* 2007) were also created by single crossover homologous recombination using pJIR1456 developed by O'Connor *et al.* (2006).

Most significantly *tcdA* and *tcdB* were inactivated using the O'Connor *et al.* (2006) homologous recombination system, leading to the suggestion that toxin B is required for virulence (Lyras *et al.* 2009).

However, as Heap *et al.* (2009a) points out, the main issue with the O'Connor *et al.* (2006) system is that as it uses single homologous recombination, mutants do not appear to be stable and removal of any selection returns the wildtype gene (as found by Dineen *et al.* (2007)). Secondly, that both the pMTL31 (which contains a Gram negative ColE 1 -derived replication origin) and the pJIR1456 plasmid (based on pIP404 of *Clostridium perfringens*) were shown to replicate poorly in *C. difficile*, possibly due to the host RM system (Purdy *et al.* 2002). Purdy *et al.* (2002) found that by utilising the pCD6 plasmid replicon found on the indigenous *C. difficile* plasmid, using a strain cured of the pCD6 plasmid (for erythromycin resistance) and using an *E. coli* strain (CA434) for conjugation, which contains modification methylase Sau96I of *Staphylococcus aureus* (M.Sau96I), it was possible to circumvent host RM systems (Purdy *et al.* 2002). The result was the ability to introduce autonomously replicating vectors into *C. difficile*.

5.1.1.1 The ClosTron

The adaptation of a novel stable gene interruption system using retargeted group II introns, the ClosTron (Heap *et al.* 2007; Heap *et al.* 2010), has enabled the almost routine insertional inactivation of *C. difficile* genes and greatly enhanced the molecular tools to study *C. difficile* pathogenesis.

Group II introns, first described in 1986 (Peebles *et al.* 1986), are large autocatalytic RNAs found in plants, eukaryotes and bacterial genomes (Toro *et al.* 2007). The ClosTron system uses a group II intron from *Lactococcus lactis* (Ll.ltrB), which after identifying its target site by base-pairing (between the excised intron lariat RNA and the target site DNA), inserts into the gene via an RNA-mediated 'retrohoming' mechanism (Lambowitz & Zimmerly 2004) (Figure 5.1.1).

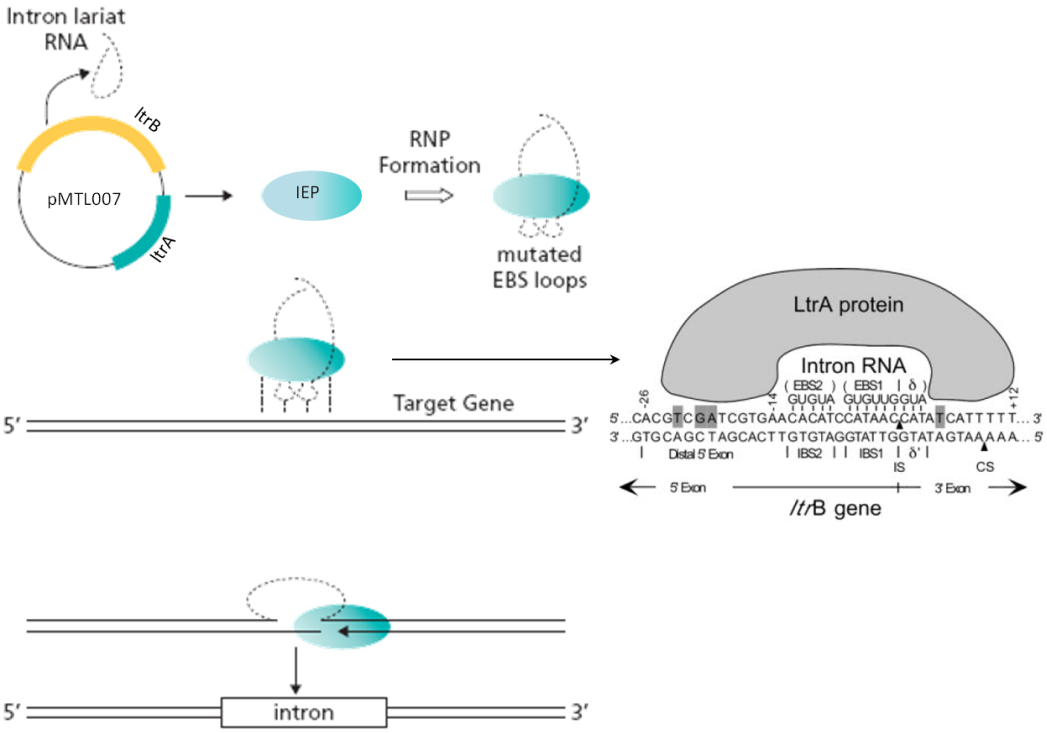


Figure 5.1.1 Mechanism of group II intron insertion. Following transcription of the LtrA protein (or intron encoded protein IEP) and the ltrB lariat RNA, they assemble to form a ribonucleoprotein (RNP). The IEP first recognises a few bases flanking regions of the DNA target site and facilitates DNA unwinding enabling the EBS1, EBS2 and δ sequences in the intron RNA to base pair with the IBS1, IBS2 and δ sequences in the DNA target site (positions -12 to +3). The ltrB then reverse-splices into the sense strand at insertion site (IS). The LtrA protein then cleaves the second (antisense) strand downstream of the intron insertion site (second strand cleavage site (CS)), providing a template for reverse transcription by the LtrA protein. Thereby leaving the target gene interrupted with a stable copy of *ltrB*. Adapted from Zhong *et al.* (2003), Karberg *et al.* (2001) and the Sigma TargeTron Manual.

As the intron target site is base-pair determined, it is reprogrammable i.e. can be mutated, to effectively target any gene (Guo *et al.* 2000; Karberg *et al.* 2001). Algorithms are used, firstly to find optimal insertion sites for the intron i.e. a best match to the positions recognised by the IEP, and secondly, to design primers to mutate the intron RNA's EBS1, 2 and δ sequences, and because the intron is

plasmid encoded, the IBS is made complementary to the retargeted EBS sequences for efficient precursor RNA splicing (Perutka *et al.* 2004).

The insertion is permanent as the LtrA (IEP) region of Ll.LtrB which harbours reverse transcriptase, maturase and endonuclease activity (Matsuura *et al.* 1997), is located on a plasmid, pMTL007 (Figure 5.1.2). Removal of the plasmid, removes the ability of the intron to transpose into another gene. Integrants are selected by use of a Retrotransposition-Activated Marker (RAM) (Zhong *et al.* 2003). The ClosTron RAM consists of an erythromycin resistance (*ermB*) ORF interrupted by a td phage group I intron. Upon expression of the group II intron RNA, the group I intron splices out, creating coding erythromycin resistance ORF, which together with a erythromycin sensitive strain of *C. difficile* (Hussain *et al.* 2005) means integrants are easily isolated.

To date, a number of *C. difficile* proteins have been knocked out using the ClosTron system including the first *C. difficile* double, toxin A and B, mutant (Kuehne *et al.* 2010).

- Phase variable surface protein CwpV (Emerson *et al.* 2009a)
- Surface located cysteine protease Cwp84 (Kirby *et al.* 2009)
- Fibronectin binding protein Fbp68 (Lin *et al.* 2010)
- The flagella core subunit FliC (Twine *et al.* 2009)
- Sporulation pathway associated protein Spo0A and SleC (Underwood *et al.* 2009; Burns *et al.* 2010b)
- Transcriptional control CodY (which may have surface based responses) (Dineen *et al.* 2010)
- FliC and FliD (Baban *et al.* 2010)

The ClosTron system is therefore a powerful tool to study the pathogenesis of the CDI via a negative phenotype approach.

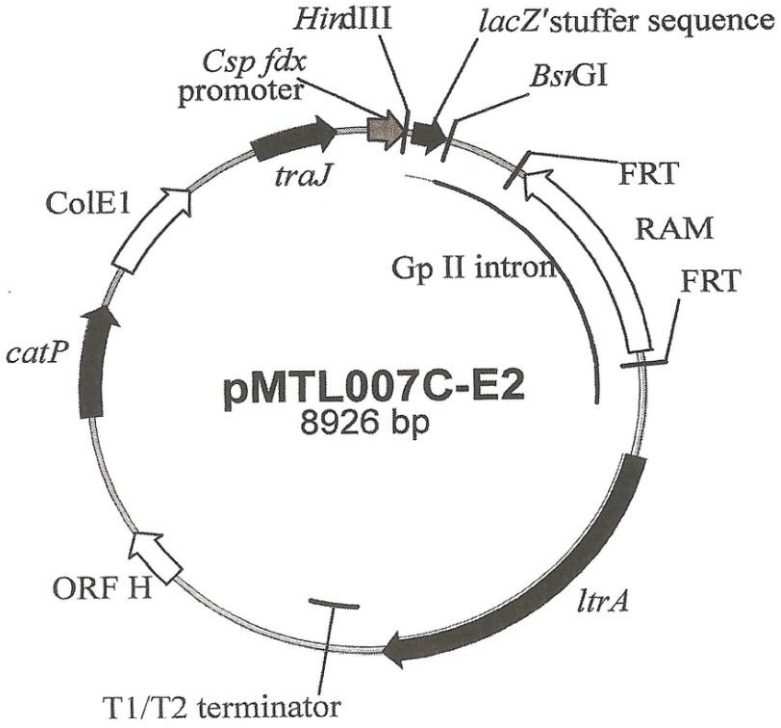


Figure 5.1.2 The ClosTron plasmid pMTL007C-E2. The *fdx* promoter of C.sporogenes drives expression of the group II intron, which contains FRT sites flanking the ErmRAM allowing marker removal. The *lacZ* stuffer is replaced during intron retargeting. Vector map taken from Heap *et al.* (2009a).

5.1.2 Surface proteins at targets for gene knockout

A total of 28 *slpA* paralogs have been identified in *C. difficile* genome (Calabi *et al.* 2001; Karjalainen *et al.* 2001; Sebaihia *et al.* 2006). All of these proteins (including *slpA*) have at their N- or C-terminus, two or three repeat regions with homology to the N-acetylmuramoyl-L-alanine amidase CwlB of *Bacillus subtilis*. This so called cell wall binding domain 2 (Pfam 04122) (CWBD) putatively mediates binding of the protein to the cell surface by non-covalent interactions. The non-covalent nature of this interaction is derived from the fact that the S-layer can be removed from the cell surface using EDTA, low pH or chaotroptic agents.

Twelve (inc. *slpA*) of the CWBD containing proteins exist in a 37 kbp cluster surrounding *slpA* and, as such, this region has been the focus of initial research efforts. From this cluster (other than *slpA*) two proteins have been characterised:

Cwp66 (CD2789) (Waligora *et al.* 2001) a heat shock inducible adhesin (See Chapter 4 Section 6.1.2.2) and Cwp84 (CD2787) a surface located cysteine protease (Janoir *et al.* 2007) (See Chapter 3). The other CWBD proteins are distributed throughout the genome, e.g. *cd0514* (*cwpV*), a phase variable cell surface protein (Emerson *et al.* 2009b).

However, despite the identification of 28 *slpA* paralogs with 7 of which identified on the cell surface (at one point or another), only 3 have been studied in any detail (Cwp66, Cwp84 and CwpV (Waligora *et al.* 1999; Janoir *et al.* 2007; Emerson *et al.* 2009b)), with an *in vivo* function only identified for one, Cwp84 (Kirby *et al.* 2009; Dang *et al.* 2010)). The role of the CWBD surface proteins in the physiology and pathogenesis of *C. difficile* has therefore only started to be understood and requires much work.

A number of these CWBD paralogs have been proven to be expressed by RT-PCR during normal growth, identified proteomically in low pH glycine extracts or lysozyme extracts (Table 5.1.1). However, the proteins found on the cell surface do not necessarily correspond to those for which a message has been found, e.g. CD2782, CD2794. These differences may be due to the type of peptidoglycan-CWBD interaction, extraction technique, extraction conditions, e.g. Cwp66's presence on the cell surface is predominantly after heat shock (Waligora *et al.* 2001), or a non-direct relationship between gene expression and protein production.

Given the number of surface proteins genes yet to have even basic characterisation, efforts were made to knock out a number of surface associated genes using the ClosTron system (Table 5.1.2). The chosen targets represent; varied genomic loci, characterised and uncharacterised surface proteins, surface proteins with either N- or C- terminal CWBDs and proteins with predicted enzymatic function and those with none.

The specific approach taken was derived from studies in Chapter 3, whereby intron insertion at any site in a gene appeared to result in the same phenotype. As such,

only one target site was attempted for each gene. However, using this approach, not every gene chosen was successfully knocked out, decreasing both the number and range of genes investigated. Only successful mutants and the associated genes are discussed further.

Table 5.1.2 Surface associated genes selected for knockout.

CWBD containing		Cell envelope associated	Pseudogenes
slpA	*cd2791 (cwp2)*	*fliD*	*cd3008*
cd1036 (cwp17)	*cwp66*	*cd2797*	*cd3146*
cd1751 (cwp13)	*cwp84 (chapter1)*		*cd2605*
cd2713 (cwp22)	*cd2767 (cwp19)*	Only genes in blue were successfully knocked out. Newly proposed gene name in brackets (Fagan *et al.* 2011).	
cd2735 (cwp14)	*cd2795 (cwp11)*		
cd2784 (cwp6)			

Table 5.1.1 All 29 CWBD containing proteins of *C. difficile*

CWBD containing protein (Alternative name)*	Gene expressed during normal growth	Protein found in glycine extracts	Protein found in cell wall extracts
CD0440 (Cwp27)			
CD0514 (CwpV)		X[3,6]	X[3]
CD0844 (Cwp25)		X[3]	X[3]
CD1035 (Cwp16)			
CD1036 (Cwp17)			
CD1047 (Cwp18)			
CD1233 (Cwp26)			
CD1469 (Cwp20)			
CD1751 (Cwp13)			
CD1803 (Cwp23)			
CD1987 (Cwp28)			
CD2193 (Cwp24)		X[3]	
CD2518 (Cwp29)			
CD2713 (Cwp22)			
CD2735 (Cwp14)			
CD2767 (Cwp19)			
CD2782 (Cwp7)	X[1]		
CD2784 (Cwp6)	X[1]	X[3]	
CD2786 (Cwp5)	X[1]		
CD2787 (Cwp84, Cwp4)	X[1]	X[4]	X[3]
CD2789 (Cwp66, Cwp3)	X[1]		X[5]
CD2791 (Cwp2)	X[1]	X[3]	X[3]
CD2793 (SlpA)	**X[1]**	**X[3]**	**X[3]**
CD2794 (Cwp12)	X[2]		
CD2795 (Cwp11)			
CD2796 (Cwp10)			
CD2798 (Cwp9)			
CD2799 (Cwp8)			
CD3192 (Cwp21)			

[1](Calabi *et al.* 2001) [2](Savariau-Lacomme *et al.* 2003) [3](Wright *et al.* 2005) [4](Janoir *et al.* 2007) [5](Waligora *et al.* 2001) [6](Emerson *et al.* 2009b). *(CwpXX) denote newly proposed nomenclature of C. difficile Pfam 04122 containing cell wall proteins (Fagan *et al.* 2011)

5.1.3 Surface proteins knocked out

5.1.3.1 CD2791 and Cwp66

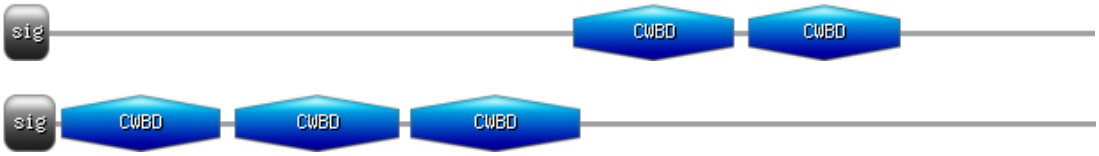


Figure 5.1.3 Domain structure of CD2791 (Top) and Cwp66 (bottom). Sig- Signal Peptide, CWBD- Cell wall binding domain.

Both CD2791 and Cwp66 are discussed in Chapter 4 Section 6.1.2.1 (CD2791) and 6.1.2.2 (Cwp66). Briefly, CD2791 is an immunogenic surface protein, its expression has been demonstrated, as has its existence on the cell surface. CD2791 is highly conserved, as such, antibodies to CD2791 cross react across strains (Calabi & Fairweather 2002; Wright *et al.* 2005; Wright *et al.* 2008). The role of CD2791 in CDI or bacterial physiology is unknown.

Cwp66 is a heat shock inducible adhesin, i.e. it is only found on the cell surface after heat shock. Antibodies to Cwp66 inhibit adhesin of *C. difficile* to Vero cells suggesting Cwp66 plays a role in adhesion (Waligora *et al.* 2001). The C-terminal Cwp66 is variable thus Cwp66 is suggested to be an important adhesin requiring antigenic variation (Lemee *et al.* 2005). Moreover, CD2791 and Cwp66 are putatively found on the same (polycistronic) transcript (Savariau-Lacomme *et al.* 2003).

5.1.3.2 CD1036 and CD2784

Table 5.1.3 Comparison of domain structures of CD1036 and CD2784

CD2784	**CD1036**
675 aa (MW 73024)	677 aa (MW 74195)
Signal peptide 1-28	Signal peptide 1-32
3 x CWBD (170-261,271-364,377-456)	3 x CWBD (174-265, 275-368, 381-474)
N-acetylmuramoyl-L-alanine amidase (490-671)	N-acetylmuramoyl-L-alanine amidase (492-673)
Active residues (H497,E514, H567, E640)	Active residues (H499,E513, H566, E642)
Metal binding (H497,E514, H567)	Metal binding (H499,E513, H566)

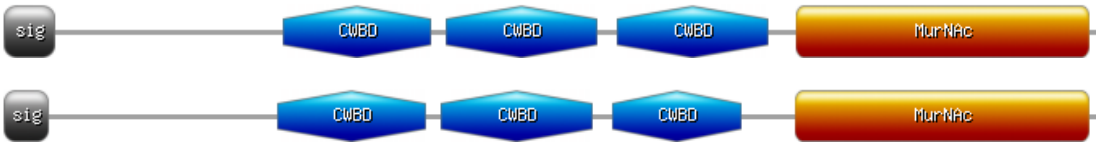


Figure 5.1.4 Domain structure of CD1036 (Top) and CD2784 (Bottom). Sig - Signal peptide, MurNAc-LAA - N-acetylmuramoyl-L-alanine amidase, CWBD – Cell wall binding domain (PFam04122).

RT-PCR revealed that *cd2784* (*cwp6* aka ORF12) was expressed during normal growth and CD2784 has been found on the cell surface in both glycine extracts and lysozyme extracts, suggesting that it is surface located and found under normal *in vitro* growth conditions (Calabi *et al.* 2001; Karjalainen *et al.* 2001; Wright *et al.* 2005). Microarray analysis found at least one reporter of CD2784 was upregulated during sub MIC amoxicillin (Emerson *et al.* 2008).

CD2784 has its CWBDs at the N-terminus and a metal binding N-acetylmuramoyl-L-alanine amidase (or MurNAc-LAA)(also known as peptidoglycan amidase) at its C-terminal (Figure 5.1.4). N-Acetylmuramyl-L-alanine amidases hydrolyse the amide bond between N-Acetylmuramic acid (MurNAc) and L-alanine (L-Ala), separating the glycan strand from the peptide (Figure 5.1.5). These proteins are Zn-dependent

peptidases with highly conserved residues involved in cation co-ordination. They are diverse in their roles including sporulation and germination, separation of daughter cells, lysis of prey cells, autolysis and peptidoglycan recycling amongst others (Vollmer *et al.* 2008).

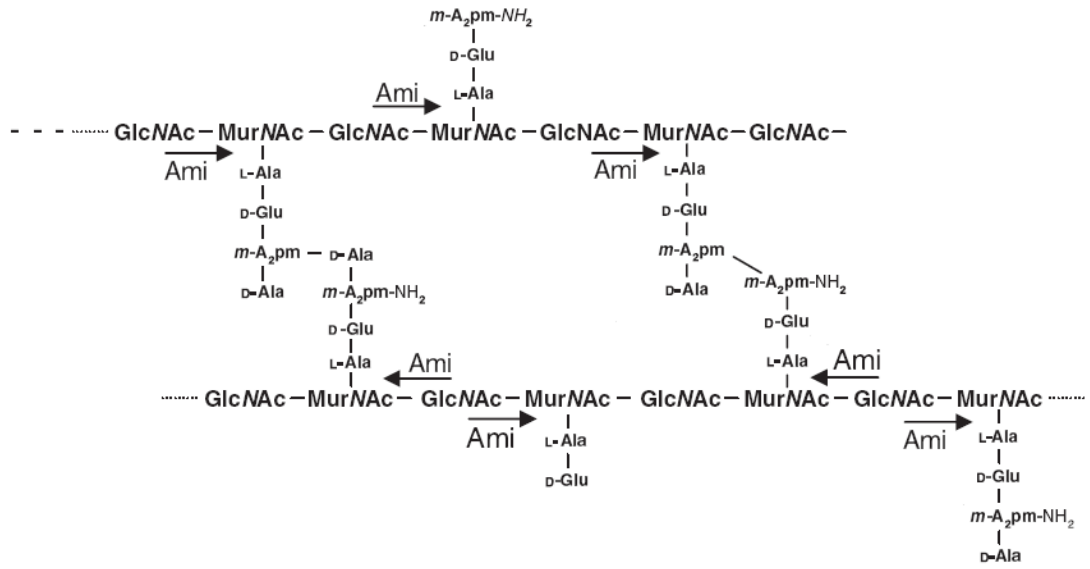


Figure 5.1.5 Hydrolysis of amide bonds in peptidoglycan**.** N-acetylmuramyl-L-alanine amidases (Ami) hydrolyse the amide bonds between the lactyl group of MurNAc and the L-alanine of the stem peptide. Adapted from (Vollmer *et al.* 2008).

C. difficile, in common with *E. coli* has several peptidoglycan amidases (CD2784, CD1035, CD1036, CD2761, CwlD (CD0106), CD2184, CD0784 and several phage endolysins (CD0972, CD1898)) located in diverse locations in the genome. Their diverse locations, suggests diverse roles in cellular physiology.

CD1036 (and CD1035 which shares 94.3% identity) is located upstream of 23s rRNA and downstream of a helicase and ATP- dependant nucleases in the *C. difficile* 630 genome. To date, no empirical studies have been performed characterising either *cd1036* or *cd1035*, including expression or the cell surface location. It is interesting that both CD2784 and CD1036/1035 share similar domain architectures but are in completely different genomic locations, this may suggest diversity of roles or gene duplication. Indeed the identification of CD2784 on the cell surface and not CD1036 suggests they are under different regulatory systems.

5.1.3.3 CD2735

Signal peptide ……1-27

3 x CWBD ……. 39-125, 134-227, 232-319

Bacterial SH3 domain ……. 346-398, 418-470

Figure 5.1.6 Domain structure of CD2735. Sig - Signal peptide, CWBD – Cell wall binding domain (PFam04122), SH3 – Bacterial Src Homology 3 domain.

The *cd2735* gene is located upstream of mleN (a putative malate-2H(+)/Na(+)-lactate antiporter, involved in Na+ and alkali resistance (Wei *et al.* 2000)) and downstream of CD2736 (a LysR family transcriptional regulator, involved in activation of divergent transcription of linked target genes or unlinked regulons encoding extremely diverse functions (Schell 1993)) in the 630 genome.

CD2735 has not been characterised to date but contains two Src Homology 3 (SH3) domains (Figure 5.1.6). SH3 domains are involved in protein-protein interactions mediating the assembly of large multi-protein complexes (Li 2005). They contain a proline-rich (PxxP) core-conserved binding motif, first identified as a conserved sequence in the non-catalytic part of several cytoplasmic protein tyrosine kinases, e.g. Src. The surface of the SH3-domain bears a flat, hydrophobic ligand-binding pocket which consists of three shallow grooves defined by conservative aromatic residues in which the ligand adopts an extended left-handed helical arrangement (Nguyen *et al.* 1998). The affinity of SH3 domains for their ligand is relatively low (K_d in µM range) and selectivity is marginal at best, but either through multiple interactions or tertiary compartmentalised interaction(s), affinity or specificity can be increased (Mayer 2001).

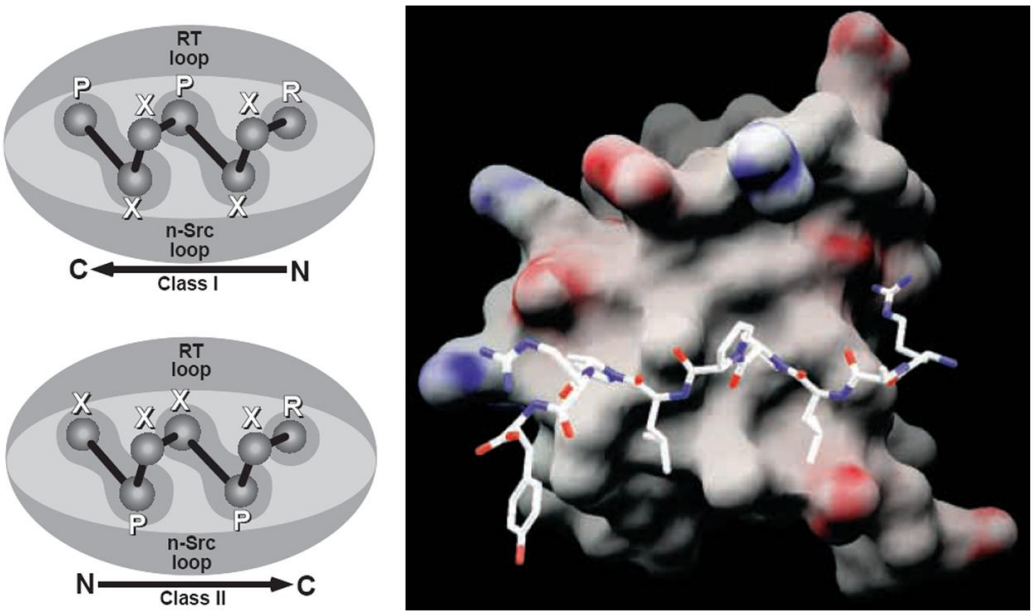


Figure 5.1.7 The structure of the two classes of SH3 domain (Feng *et al.* 1994) and surface of the SH3 domain with ligand. Figure adapted from Mayer (2001).

Inhibition of the SH3 domain can be accomplished by synthetic proline rich ligands (Lee *et al.* 2002b). Inhibition of the SH3 protein-protein interactions and the subsequent inhibition of the commonly associated (kinase) signalling appear to be key in controlling tumorigenesis (Lu *et al.* 2010).

The role of SH3 domains in bacterial pathogenesis is largely unknown but SH3 domains are present in *Listeria monocytogenes* invasion protein InlB, mediating binding to host ligands (Marino *et al.* 2002) and iron regulation in Mycobacterium sp (Liu *et al.* 2008a).

5.1.3.4 CD2795 cwp11 (Calabi *et al.* 2001) ORF5 (Karjalainen *et al.* 2001)

533 aa (58,743Da predicted)

Signal peptide 1-28

3 x CWBD 31-118, 127-213, 217-296)

Bacterial Ig-like domain (group 3); pfam07523 309-361

SCP_bacterial 398-525

Figure 5.1.8 Domain structure of CD2795 (and CD2794). Sig - Signal peptide, CWBD – Cell wall binding domain (PFam04122), Big_3 - Bacterial Ig-like domain (group 3)(PFam07523), SCP_bacterial - SCP-like extracellular protein domain

In the *C. difficile* 630 genome, *cd2795* is upstream of *cd2794*, which is upstream of *slpA*. Both CD2795 and CD2794 share a similar domain arrangement (Figure 5.1.8), although CD2794 is 5 aa shorter (528 aa). However, there is only 70.3% identity at the DNA level and 63.4% at the amino acid level between *cd2794* and *cd2795*. Despite being in the *slpA* cluster (indeed next to *slpA*), no characterisation has been undertaken with *cd2795* nor has CD2795 been identified on the cell surface, suggesting that either it is not expressed, or its expression is condition dependant (although *cd2794* has been proven to be expressed by RT-PCR (Karjalainen *et al.* 2001)).

CD2795 contains a bacterial Ig-like domain (group 3) (PF07523) which contains an Ig-like fold and is putatively found in a range of bacterial surface proteins. The bacterial Ig-like domain has been found in the *E. coli* adhesion protein intimin, in which two Ig-like domains are followed C-terminally by a lectin binding domain (Kelly *et al.* 1999; Luo *et al.* 2000). An immunoglobulin-like fold consists of a beta-sandwich of seven or more strands in two sheets with a Greek-key topology (Bork

et al. 1994). Ig-like domains often interact with other Ig-like folds, via their beta-sheets (Potapov *et al.* 2004).

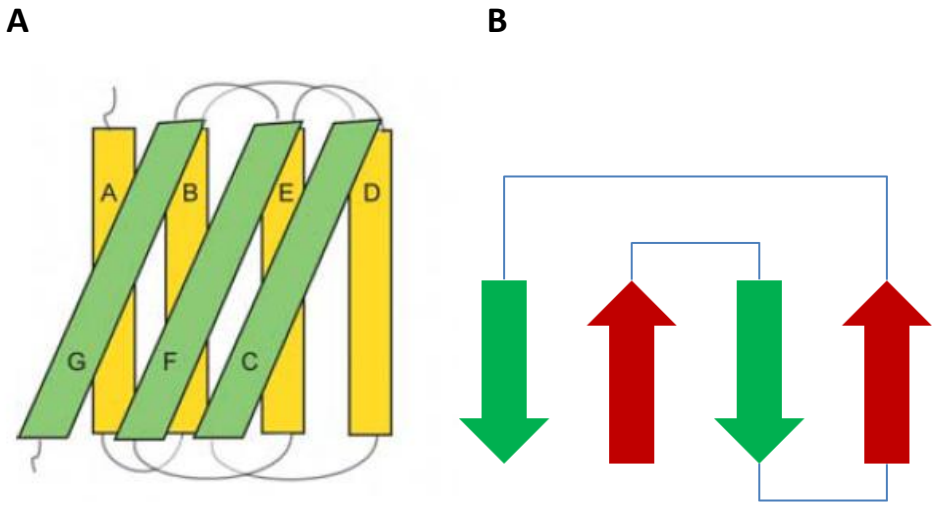


Figure 5.1.9 Schematic presentation of the immunoglobulin like fold (A) and the Greek-key topology (B).

It is interesting that two proteins (CD2794 and CD2795), containing the Ig-like domain, are found next to each other in the *C. difficile* genome, both with CWBD. It is tempting to speculate that they interact possibly forming a protein complex on the cell surface, or bind directly to host molecules forming a complex.

At the C-terminal of CD2795 is a SCP bacterial domain. The sperm coat protein (SCP) bacterial domain has been found in plants (PR1-like proteins), cysteine-rich secretory proteins (CRISPs), reptiles and insects (ion channel blocking venoms) and nematodes (Yeats *et al.* 2003). Certain SCP proteins contain a calcium chelating serine protease active site (Fernandez *et al.* 1997), while others do not, but contain conserved residues including histidine and glutamate which may co-ordinate ligand(s), for example zinc in *Pseudecin* snake venom (Suzuki *et al.* 2008), however the role of the SCP domain in bacterial pathogenesis is unknown.

5.1.3.5 FliD

507 aa (55,946Da)

FliD (Flagella capping protein) ……. ……8-506

FliD_N (Flagellar hook-associated protein 2 C-terminus).....13-117

FliD_C (Flagellar hook-associated protein 2 C-terminus).....233-496

Figure 5.1.10 Domain structure of FliD - Flagella cap protein. FliD_N - Flagellar hook-associated protein 2 C-terminus (PFam 02465). FliD_C - Flagellar hook-associated protein 2 C-terminus (PFam07195).

At its simplest level, flagella are used to move the bacterium from one place to another. This locomotion provides the bacterium access to nutrients via chemical gradients. The motile flagellum consists of a number of elements, assembled in a particular order, topped off with a cap (FliD) (Macnab 2003) (Figure 5.1.11)

The expression of the flagella cap (*fliD*) gene in *C. difficile* was found in flagellated and non-flagellated strains. However, anti-FliD serum did not react with a 'flagella extract' of non-flagellated, strains suggesting gene expression and protein production are not necessarily correlated. FliD appears to be highly conserved and falls into two major forms with 88% identity between them (Tasteyre *et al.* 2001b). However, despite the conservancy, Lemee et al (2005) suggests *fliD* (and *fliC*) are under selective pressure and demonstrate a number of polymorphisms. Flagellated *C. difficile* strains adhered to Vero cells tenfold greater than non-flagellated strains. Furthermore, recombinant FliD binds to mucus and radiolabelled Vero cells bind to recombinant FliD respectively (Tasteyre *et al.* 2001a), supporting a role for the flagella in bacterial adhesion.

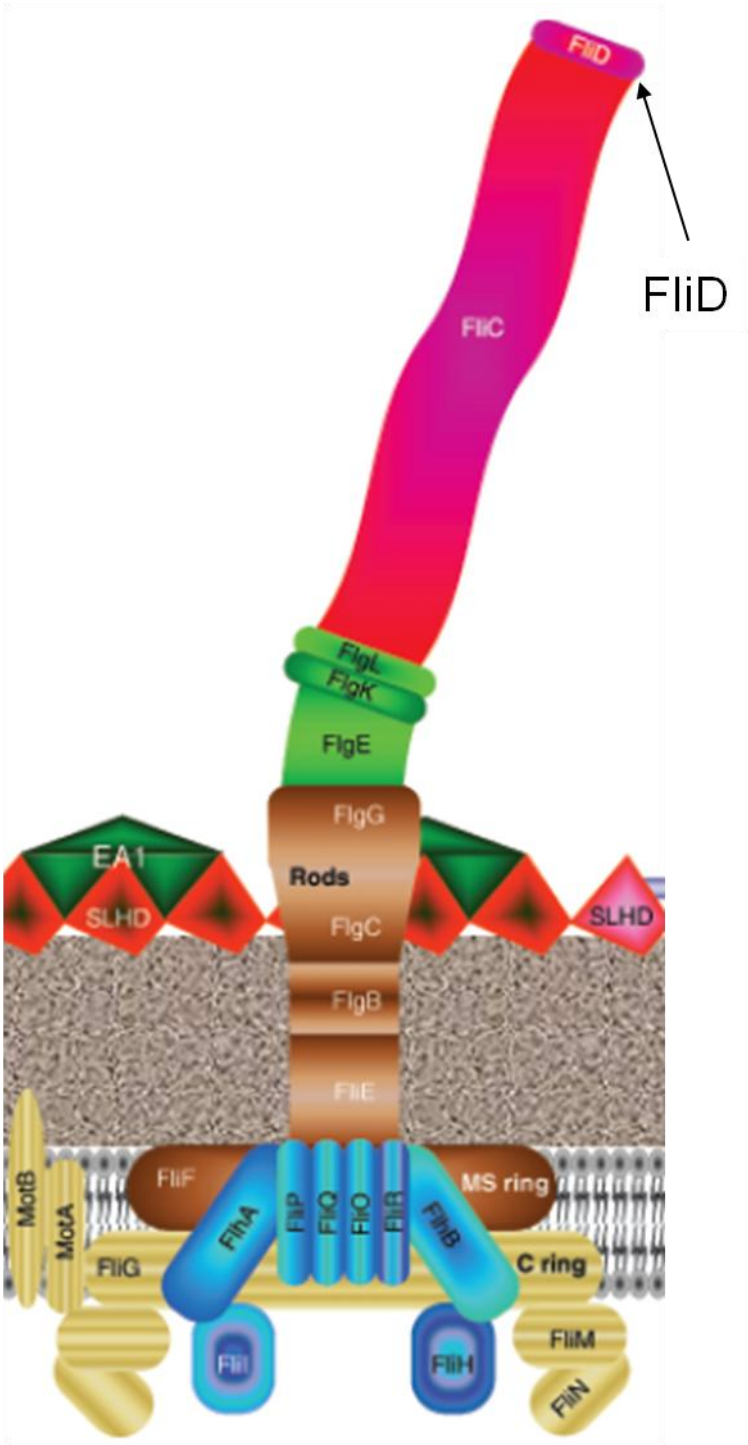


Figure 5.1.11 Schematic representation of the Gram positive flagella. Taken from (Desvaux *et al.* 2006).

Consistent with the role of the flagella in *C. difficile* pathogenesis, the majority of patients appear to raise antibodies to FliD (Pechine *et al.* 2005b) and responses appear to be to the same level as toxin A and B, possibly higher in the control group than the CDI case group (Pechine *et al.* 2005b). Rectally administered encapsulated

FliD plus a flagella extract, raised an anti-colonisation protective response in mice (Pechine *et al.* 2007), suggesting FliD plays a role in colonisation of the gut.

5.1.4 Aim

The aim of this study was to understand more about selected surface proteins using a gene knockout approach, followed by characterisation of mutants to identify surface based determinants of virulence worth further investigation. This chapter has determined that inactivation of several surface proteins results in changes in the bacterium, which both result in a decreased ability to adhere *in vitro* and appears to cause an earlier release of toxin compared to the wildtype. However, other aspects of cell physiology e.g. sporulation, morphology and removed surface proteins (beyond the species knocked out) are the same as the wildtype.

5.2 Chapter 5 Specific Methods

5.2.1 Knockout of putative surface protein genes

The ClosTron gene knock-out system was used for insertional inactivation using protocols outlined in General Methods Section 2.2.

Intron (re)targeting region using primers designed by the relevant intron retargeting site are given in Table 5.2.1. Primers used to screen for intron integration are given in Table 5.2.2.

Table 5.2.1 Intron retargeting primers, as designed by ClosTron and Sigma TargeTron algorithms

Gene	Insertion location	Gene-intron-gene	Retargeting primers		
			IBS	IBS1d	EBS2
cd1036	1406|1407a	TTGTGTTTAGATAATGATGATGCTATAGAG- intron -CTCATTACCTTAGAA	AAAAAAGCTTATAATTATCCTTAGATGCCATAGAGGTGCGCCCAGATAGGGTG	CAGATTGTACAAATGTGGTGATAACAGATAAGTCATAGAGCTTAACTTACCTTTCTTTGT	TGAACGCAAGTTTCTAATTTCGATTGCATCTCGATAGAGGAAAGTGTCT
cd2735	1113|1114s	GGATATCTAAATAATGGTGATGAAGTAGAA- intron -GTTTTAGATGTGTTA	AAAAAAGCTTATAATTATCCTTAGATGACGTAGAAGTGCGCCCAGATAGGGTG	CAGATTGTACAAATGTGGTGATAACAGATAAGTCGTAGAAGTTAACTTACCTTTCTTTGT	TGAACGCAAGTTTCTAATTTCGATTTCATCTCGATAGAGGAAAGTGTCT
cd2784	438|439s	AATCTTAAGGATTATGATGTTACACAAGAG- intron -TTTATAATAAGTGAT	AAAAAAGCTTATAATTATCCTTAGTTACCCAAGAGGTGCGCCCAGATAGGGTG	CAGATTGTACAAATGTGGTGATAACAGATAAGTCCAAGAGTTTAACTTACCTTTCTTTGT	TGAACGCAAGTTTCTAATTTCGATTGTAACTCGATAGAGGAAAGTGTCT
cd2791	1569|1570s	GCTGATGTTGATAAAGATAGAAAAGTTCAA - intron -AGAGTTGAAGGAGAA	AAAAAAGCTTATAATTATCCTTAAGAAACGTTCAAGTGCGCCCAGATAGGGTG	CAGATTGTACAAATGTGGTGATAACAGATAAGTCGTTCAAAGTAACTTACCTTTCTTTGT	TGAACGCAAGTTTCTAATTTCGATTTTTCTTCGATAGAGGAAAGTGTCT

cd2795	1320|1321s	TATATGATG GATAAAAAA GTATTTGCAC AT- intron - TATATAGATG GAAAA	AAAAAAGCT TATAATTATC CTTAGTATTC GCACATGTG CGCCCAGAT AGGGTG	CAGATTGTAC AAATGTGGT GATAACAGA TAAGTCGCA CATTATAACT TACCTTTCTT TGT	TGAACGCAA GTTTCTAATT TCGATTAATA CTCGATAGA GGAAAGTGT CT
cwp66	519|520s	AAGGGTCTT GCTGATGCA GTGAGTGTT GGA - intron - GCTATAGCT GCTCAA	AAAAAAGCT TATAATTATC CTTAGTGAG CGTTGGAGT GCGCCCAGA TAGGGTG	CAGATTGTAC AAATGTGGT GATAACAGA TAAGTCGTTG GAGCTAACTT ACCTTTCTTT GT	TGAACGCAA GTTTCTAATT TCGATTCTCA CTCGATAGA GGAAAGTGT CT
cwp84	347|348a	TCTAAAGTTG ACATACCTGA AAAAGACCA G - intron - CATGTATTAA GACTT	AAAAAAGCT TATAATTATC CTTAGAAAA CGACCAGGT GCGCCCAGA TAGGGTG	CAGATTGTAC AAATGTGGT GATAACAGA TAAGTCGAC CAGCATAACT TACCTTTCTT TGT	TGAACGCAA GTTTCTAATT TCGATTTTTT CTCGATAGA GGAAAGTGT CT
fliD	120|121s	GAAAAAGTT GATAAAGCA AAACAAGAA CAA - intron - CAAATCGTTA AATGG	AAAAAAGCT TATAATTATC CTTAAAACAC GAACAAGTG CGCCCAGAT AGGGTG	CAGATTGTAC AAATGTGGT GATAACAGA TAAGTCGAA CAACATAACT TACCTTTCTT TGT	TGAACGCAA GTTTCTAATT TCGGTTTGTT TCCGATAGA GGAAAGTGT CT
cd1751	2024|2025a	AGATTGTTTG ATAGGGATG CTGTTCCTCC A- intron - ATAATATAA GATTTC	AAAAAAGCT TATAATTATC CTTAGCTGTC CCTCCAGTGC GCCCAGATA GGGTG	CAGATTGTAC AAATGTGGT GATAACAGA TAAGTCCCTC CAATTAACTT ACCTTTCTTT GT	TGAACGCAA GTTTCTAATT TCGATTACAG CTCGATAGA GGAAAGTGT CT
cd2713	1482|1483a	TGGTTATTTG ACCAAGTTCC TGTAGGTTCA - intron - GATGTAATTA TAGAT	AAAAAAGCT TATAATTATC CTTACCTGTC GGTTCAGTG CGCCCAGAT AGGGTG	CAGATTGTAC AAATGTGGT GATAACAGA TAAGTCGGTT CAGATAACTT ACCTTTCTTT GT	TGAACGCAA GTTTCTAATT TCGGTTACA GGTCGATAG AGGAAAGTG TCT
cd2767	1524|1525a	TACGATATA GATGGAGTT CATTTTGATG AT- intron - TACTTCTATC CAGGA	AAAAAAGCT TATAATTATC CTTACATTTC GATGATGTG CGCCCAGAT AGGGTG	CAGATTGTAC AAATGTGGT GATAACAGA TAAGTCGAT GATTATAACT TACCTTTCTT TGT	TGAACGCAA GTTTCTAATT TCGGTTAAAT GTCGATAGA GGAAAGTGT CT

cd2797	2925|2926a	GGAGTAGCT GCTAGAGCA GGAGATGAA ATC- intron - ACAGTATAT GTTGGA	AAAAAAGCT TATAATTATC CTTAGGAGA CGAAATCGT GCGCCCAGA TAGGGTG	CAGATTGTAC AAATGTGGT GATAACAGA TAAGTCGAA ATCACTAACT TACCTTTCTT TGT	TGAACGCAA GTTTCTAATT TCGATTTCTC CTCGATAGA GGAAAGTGT CT
cd3008 (pseudo)	39|40a	TTTAATGTGG CTAGAGGTA TAATTGAAG AG- intron - TATATTAAAG AATGT	AAAAAAGCT TATAATTATC CTTAATAATC GAAGAGGTG CGCCCAGAT AGGGTG	CAGATTGTAC AAATGTGGT GATAACAGA TAAGTCGAA GAGTATAAC TTACCTTTCT TTGT	TGAACGCAA GTTTCTAATT TCGGTTATTA TCCGATAGA GGAAAGTGT CT
cd3146 (pseudo-sortase)	120|121a	GCTATTATTG ATAGTGTTTT CTCTGAGTTC - intron - CGTATATACT TCATC	AAAAAAGCT TATAATTATC CTTATTCTCC GAGTTCGTG CGCCCAGAT AGGGTG	CAGATTGTAC AAATGTGGT GATAACAGA TAAGTCGAG TTCCGTAACT TACCTTTCTT TGT	TGAACGCAA GTTTCTAATT TCGGTTGAG AATCGATAG AGGAAAGTG TCT
cdtB (CD2605) (pseudo)	645|646s	ATACCAATA GATGAAAGT TGTGTTGAAC TC- intron - ATATTTGATG ATAAT	AAAAAAGCT TATAATTATC CTTATGTGTC GAACTCGTG CGCCCAGAT AGGGTG	CAGATTGTAC AAATGTGGT GATAACAGA TAAGTCGAA CTCATTAACT TACCTTTCTT TGT	TGAACGCAA GTTTCTAATT TCGGTTACAC ATCGATAGA GGAAAGTGT CT
slpA	1266|1267s	GCAATAACT GATAAAGCA GTTAATGATA TA - intron - GTATTAGTTG GATCT	AAAAAAGCT TATAATTATC CTTAGTTAAC GATATAGTG CGCCCAGAT AGGGTG	CAGATTGTAC AAATGTGGT GATAACAGA TAAGTCGAT ATAGTTAACT TACCTTTCTT TGT	TGAACGCAA GTTTCTAATT TCGATTTTAA CTCGATAGA GGAAAGTGT CT
slpA	258|259s	AAGAAAGCG GACAGAGAT GCTGCAGCT GAG - intron - AAGTTATATA ATCTT	AAAAAAGCT TATAATTATC CTTAGCTGCC GCTGAGGTG CGCCCAGAT AGGGTG	CAGATTGTAC AAATGTGGT GATAACAGA TAAGTCGCT GAGAATAAC TTACCTTTCT TTGT	TGAACGCAA GTTTCTAATT TCGATTGCA GCTCGATAG AGGAAAGTG TCT
slpA	321|322s	AAATTAGGT GATGGAGAT TATGTTGATT TT - intron - TCTGTAGATT ATAAT	AAAAAAGCT TATAATTATC CTTATATGTC GATTTTGTGC GCCCAGATA GGGTG	CAGATTGTAC AAATGTGGT GATAACAGA TAAGTCGATT TTTCTAACTT ACCTTTCTTT GT	TGAACGCAA GTTTCTAATT TCGGTTACAT ATCGATAGA GGAAAGTGT CT

tcdR	499|500s	AAATATCTAGACAAGCTGTTAATAAGGCTA- intron -AAAATAGAGCATTTA	AAAAAAGCTTATAATTATCCTTATTAATCAGGCTAGTGCGCCCAGATAGGGTG	CAGATTGTACAAATGTGGTGATAACAGATAAGTCAGGCTAAATAACTTACCTTTCTTTGT	TGAACGCAAGTTTCTAATTTCGGTTATTAATCGATAGAGGAAAGTGTCT
tcdA	494|495a	TTAATTGTTGATCAAGTTTATCAAAGTTTT- intron -CTGATGTATAATTAT	AAAAAAGCTTATAATTATCCTTATATCACAGTTTTGTGCGCCCAGATAGGGTG	CAGATTGTACAAATGTGGTGATAACAGATAAGTCAGTTTTCTTAACTTACCTTTCTTTGT	TGAACGCAAGTTTCTAATTTCGGTTTGATATCGATAGAGGAAAGTGTCT
Spo0A	178|179a	ATCCATCTAGATGTGGCATTATTACATCTA- intron -GTATTAATAAGTCCG	AAAAAAGCTTATAATTATCCTTATTATTCCATCTAGTGCGCCCAGATAGGGTG	CAGATTGTACAAATGTGGTGATAACAGATAAGTCCATCTAGTTAACTTACCTTTCTTTGT	TGAACGCAAGTTTCTAATTTCGGTTAATAATCGATAGAGGAAAGTGTCT

5.2.2 Characterisation of mutants

Prior to any manipulation or characterisation of mutants, cultures were re-streaked onto FAA plate(s) and incubated at 37 °C for 24 hr. Where several characterisations were performed simultaneously, the same re-streak was used.

5.2.2.1 Molecular Methods

gDNA extraction and RNA extraction was undertaken as per General Methods sections 2.1.11 and 2.1.12.2 respectively.

PCR was undertaken as per General Methods section 2.1.5, while RT-PCR was performed using either SuperScript III One-Step RT-PCR System with Platinum® Taq DNA (Invitrogen) or QIAGEN OneStep RT-PCR Kit (with Q solution) (Qiagen) using primers outlined in Table 5.2.3.

Table 5.2.2 Primers used to screen for integration

Primer	Sequence
CD2791_for	ATG AAT AAA AAA AAT CTT TCT G
CD2791_rev	TTA CCA ACC TAG TAT TTT AGC
Cwp66_For	ATG AAA ATA TCA AAA AAG ATA GTG TC
Cwp66_Rev	TAA AWT CCA TCA TCT GTA
1036 Up-279_F	GGC TAC AGT TAA TTA TCA AGA G
1036_1703_R	CCT TGT TTA TTA GTT GAG CCA T
CD2735_For(mwg)	TCA TCA ATG CTG TGC TTA TCT C
CD2735_Rev(mwg)	CCC ATC CAG TAG CAA ACA TTT C
CD2784_For	TGT TGT CTG TAG CAA TGG TG
CD2784_Rev	CTG TAT CTT CTA TTC CTA TTC C
CD2795_For	ATG AAA ATG AAC AAA AAA ATA
CD2795_Rev	TTA CAA GCT ACC TTC ATT GC
CD1751_for	GTG AAA AAA TTT ACT TC
CD1751_rev	CTA TTT AGC ACT TTT TA
CD2713_For	ATG TTT CTA AAA GAG GGG GA
CD2713_Rev	CTA TTT TGC AAG AAT ATC CA
CD2767_For	ATG AAA AAA ATA AGT ATA TT
CD2767_Rev	TTA CTT AAC TAA GTT TAA GA
CD2797_For	ATG AAA AAG GCA ATA TCT TGT
CD2797_Rev	CTA TTT ATC TCT AAC AAT AT
FliD_For	ATGTCAAGTATAAGTCC
FliD_Rev	TTAATTACCCTGTGCTT
***slpA*MasSpec_Rev (Qazi *et al.* 2009)**	ATTCTATGTACATAATAAAGAGATGT
***slpA*MasSpec_For (Qazi *et al.* 2009)**	ACCTTCACCAGTTTTCAT
tcdR_For	AAGTAAGTCTGTTTTTGAGG
tcdR_Rev	AAGGCTTTATTTCTACCAG
tcdA_For	TTTAGCTGCAGCATCTGACAT
tcdA_Rev	ATGGCTGGGTTAAGGTGTTG

Table 5.2.3 Primers used for RT-PCR

Gene	Primer name	Primers
cd1036	1036-1200PreI	AGA TGA TGA GTT AGA GGC TG
	1036-1528PostI	CTC CAG ATG AAT TTC CAG GT
cd2735	2735-1052PreI	GAA ATG ATG CTA CAA TAG AAG C
	2735-1271PostI	TTC CTA ACA TTT AGA CCA TC
cd2791	2791-1000preI	ACT GCA GTG GCT ATA GCA AA
	2791-1703PostI	CCT GCT GCT AAT GCA TCT AC
cd2791 **(Post intron)**	2791PostIn1576F	GAA GGA GAA ACA AGA CAC GA
	2791PostIn1775R	AGT GCA TTC TTT TGT GAG TC
cwp66	cwp66-200PreI	CTT TAT CAG CTA CTC CAT TT
	cwp66-525PostI	TAT AGC TCC AAC ACT CAC TG
cd2791-cd2790 **intergenic**	orf8-RT1[1]	AGT TGA ATT GAC AGT AAT AC
	orf9-RTR1[1]	CTG TGC ATA ATA TGA CAT GT
cd2790-cwp66 **intergenic**	orf9-RT1[2]	GCT ATA GGA TAT CAT TCA G
	cwp66-RTR1[2]	GAT AAA GCA TCT GCT ATG G
cd2795 **(Primer set 1)**	2795-1219PreI F	AGA AAA GAA AAA GGC AAA GAG C
	2795-1446PostIR	TGC CAA GTC TTT TGC ATC TT
cd2795 **(Primer set 2)**	CD2795_for1098	CAC TCA TTG CAT CTG CTT CTC
	CD2795_rev1478	GCA GCA AAG ATA GCT GAT AAA C

[1,2] refer to Orf8 RT1 & Orf9 RTR1 and Orf9-RT1 & cwp66 RTR1 respectively from Savariau-Lacomme *et al.* (2003).

5.2.2.2 Growth rate determination

Triplicate 20 ml sBHI broth was inoculated with 3 streaks of colonies and incubated for 16 hr. The overnight culture was inoculated into fresh triplicate 20 ml sBHI broth

such that the ~OD_{600} 0.1. At hourly intervals the culture was agitated and samples removed and OD_{600} measured. OD_{600} vs time was plotted and the linear (slope/gradient/exponential) part of the curve used to give a growth rate.

5.2.2.3 Sporulation

A stirred 80 ml sBHI culture was inoculated with a streak of colonies and left for 18 days. After this period samples were taken and plated in dilution, pre and post heat shock (62 °C, 40 min), to give viable count (colony forming units/ml).

5.2.2.4 Toxin ELISA

Sandwich ELISA for toxins A and B was performed as detailed in General Methods (Section 2.7.1).

5.2.2.5 Motility assay

The relevant *C. difficile* strain was inoculated, with a loop, into the top 2 to 5 mm of pre-reduced BHI agar (0.05%). Cultures were then incubated for 24 hr in anaerobic conditions at 37°C before being photographed.

5.2.2.6 Surface protein extraction

Surface proteins were removed by low pH glycine extraction as detailed in General Methods (Section 2.7.2).

5.2.2.7 Adhesion assay

An *in vitro* adhesion assay was performed as detailed in Chapter 4 Specific Methods (Section 4.2.6). Enumeration was performed using a single-blind experimental design.

5.2.3 Assessment of CD2791 as an adhesin

5.2.3.1 Expression and purification of $rCD2791_{27\text{-}322}$

Expression and purification of the N-terminal domain of CD2791 was performed as detailed in Chapter 4 Specific Methods (Section 6.3.1.1.2).

5.2.3.2 Labelling of $rCD2791_{27\text{-}322}$

After extensive dialysis into 50 mM HEPES plus 0.15 M NaCl pH 7.4, $rCD2791_{27\text{-}322}$ was labelled with AlexaFluor488 protein labelling kit (Molecular Probes, USA) as per

manufacturer's instructions. The A_{280} value was used to concentrate the rCD2791$_{27-322}$ to 2 mg/ml using a Vivaspin 20 10 kDa MWCO spin concentrator. Calculation of extinction co-efficient was performed with ProtParam (Gasteiger *et al.* 2005).

5.2.3.3 Binding of rCD2791$_{27-322}$ to Caco2 cells or Vero cells

Caco-2 cells were seeded into 8 well Lab-Tek™ chamber slides at a 1:2 dilution in MEME for 14 days prior to the assay to ensure full differentiation. Vero cells were seeded into chamber slides approximately 18 hr before the assay.

Cells in chamber slides were washed twice with 0.4 ml per well DPBS before addition of 200 µl of labelled or unlabelled rCD2791$_{27-322}$ or control solutions (DPBS or DMEM). Wells with 1 or 50 µM Alexa488 labelled rCD2791$_{27-322}$ were incubated for 1hour. Wells with 10 or 100 µM Unlabelled rCD2791$_{27-322}$ were incubated for 1hour, washed 3 times with 0.4 ml PBS followed by application of 1 µM Alexa488 labelled rCD2791$_{27-322}$ and incubated for a further hour. Incubations were performed at 37 °C in an atmosphere of 5% CO_2. After incubation, solutions in wells were removed followed by washing 3 times with 0.4 ml per well of DPBS before viewing on a fluorescent microscope with FITC filter. Calculation of rCD2791$_{27-322}$ µM concentrations was done by adjustment of the A_{280} using extinction co-efficient given by ProtParam (Gasteiger *et al.* 2005).

5.2.3.4 ELISA of antibody response to rCD2791$_{27-322}$

rCD2791$_{27-322}$ was coated onto Nunc Maxisorp 96-well ELISA plates at 10 µg in PBS and incubated for 1 hr. The plate was then washed 3 times with PBS/T followed by blocking PBS/T with 5% FCS. After washing as before, 1:3 serially diluted ovine anti-R20291 SLP IgG (1 mg/ml) was applied and the plate incubated for 1 hr. After washing as before, donkey anti-sheep HRP conjugate was applied (1:5000 in blocking buffer) and incubated for 1 hr. The plate(s) were washed a final time and developed with TMB substrate and TMB stop solution, then read at 450 nm using a Tecan™ Sunrise microtitre plate reader. All washes were performed 3 times with PBS plus 0.1% Tween 20 (Sigma) (PBS/T). All incubations were performed at 37 °C with shaking.

5.2.3.5 Coupling rCD2791$_{27\text{-}322}$ to NHS resin and antibody purification

5.2.3.5.1 Coupling

rCD2791$_{27\text{-}322}$ was coupled to Pierce NHS-activated agarose slurry (ThermoFisher) as per manufacturer's instructions. Briefly, 10 ml settled resin was washed three times H_20, followed by two washes with coupling/wash buffer (PBS). Approximately 17 mg rCD2791$_{27\text{-}322}$ in 50 mM HEPES 0.15 M NaCl pH 7.4 was added to the resin and incubated for 1.5 hr with end-over-end stirring. The resin was washed twice with PBS, followed by blocking with 1 M ethanolamine pH 7.6 for 20 min. Resin was again washed once with PBS before loading onto an XK16 column and washed with PBS for 100 ml (3 ml/min). The resin was stored in PBS plus 0.05% azide at 4 °C. Coupling efficiency was measured by Bradford assay, using BSA as the standard. Approximately 60% coupling was achieved.

5.2.3.5.2 Antibody purification

The 10 ml NHS.rCD2791$_{27\text{-}322}$ resin was extensively washed with tris-buffered saline (TBS) followed by application of 40 mg ovine anti-R20291 SLP IgG (1 mg/ml in TBS) at 1 ml/min. The column was washed until the OD returned to near baseline. Bound material was eluted by 100% gradient to 150 mM glycine pH 2.5, applied at 1 ml/min (1 ml fractions). Fractions were neutralised by 200 µl 1 M tris pH 9.0.

5.3 Results

5.3.1 Production of *C. difficile* surface protein knockouts

Three types of gene were chosen for knockout:

- Genes which encoded a protein containing a CWBD (Pfam 04122) repeat
- Associated with cell wall or surface located virulence factor
- Putative pseudogenes- where no functional protein is produced to assess intron insertion at a putatively unimportant site.

The following genes and insertion site(s) were chosen for investigation based on the criteria above (Table 5.3.1).

Table 5.3.1 Genes chosen for knockout (including those used in Chapter 3).

Gene	Number of predicted intron insertion sites	Site(s) chosen for investigation	Gene size (bp)	Algorithm location	KO?
cd1036	17	1406|1407a	2034	Clos	y
cd2735	14	1113|1114s	1428	Clos	y
cd2784	16	438|439s	2028	Clos	y
cd2791	15	1569|1570s	1872	TT	y
cd2795	14	1320|1321s	1602	Clos	y
cwp66	24	519|520s	1833	TT	y
cwp84	27	68|69a	2412	TT	y
cwp84	27	347|348a	2412	TT	y
cwp84	27	675|676s	2412	TT	y
cwp84	27	677|678a	2412	TT	y
cwp84	27	2054|2055a	2412	TT	y
fliD	12	120|121s	1524	TT	y
cd1751	26	2024|2025a	2388	Clos	n
cd2713	11	1482|1483a	1962	Clos	n
cd2767	10	1524|1525a	2112	Clos	n
cd2797	51	2925|2926a	5964	Clos	n
cd3008(pseudogene)	7	39|40a	714	Clos	n
cd3146(pseudosortase)	4	120|121a	690	Clos	n
cdtB(CD2605)(pseudo*)	10	645|646s	1536	Clos	n
slpA	18	1266|1267s	2160	TT	n
slpA	18	258|259s	2160	TT	n
slpA	18	321|322s	2160	TT	n

Clos- ClosTron free intron retargeting (www.clostron.com), TT- Sigma Targtron Design site retargeted.* encodes a truncated non-functional gene, therefore considered a pseudogene.

Two methods of intron retargeting were assessed: retargeting via Sigma TargeTron algorithm (http://www.sigma-genosys.com/targetron/) which is "an extensively validated TargeTron algorithm that predicts optimal intron insertion sites and designs primers for mutating the intron to insert into those sites" and the implementation of the Perutka et al (2004) algorithm available free at www.clostron.com.

The preferred intron insertion sites given by each algorithm, while containing similarities, do differ as demonstrated by the top 5 hits of Cwp84 from R20291 (Table 5.3.2).

Table 5.3.2 Differences between intron retargeting algorithm results for *cwp84* of *C. difficile* R20291, using ClosTron.com or Sigma TargeTron.

ClosTron

Position	Sequence Exon-->intron<--Exon	Score
2054|2055a	TCTAAATTTGATGATAAAGTAGCTGTACCA - CCAATTATATAAGAC	10.597
1481|1482a	CCTGAAGTAGCTTGTCCAGAAGATGTTTTT - GTTATAGCTTTTATA	8.749
2277|2278s	GCATTAGGTGCTAAAACTGATTCACCAGTA - GTTCTAGTTGGAAAT	8.697
68|69a	TGGTTTTCTGCTGAGACAGGTATTGTTGAC - ACTATTAAAAAGCAA	8.009
624|625s	TCAAATATGGATACTGCACCTACTCAATTT - AATGTAACAGATGTT	7.703

TargeTron

Position	Sequence Exon-->intron<--Exon	
481|482s	GAAAATATGGATGGAACTTAGATGATATGT - CAGGTAGTTCAAATG	11.13
1011|1012s	TCGAAACCAGATAGTGATAAAAAAATGTAC - CAACTTGAATATGCT	8.79
2054|2055a	TCTAAATTTGATGATAAAGTAGCTGTACCA - CCAATTATATAAGAC	8.73
26|27a	CAAACTGTCAACAATGCCAGTAGCTTTGAC - AATTTTTTTGATTTG	7.86
675|676s	CTTAATAAAGATAAGGAAACTGTAAAAAAT - GCTATAATGCAATAT	7.8

As discussed in Chapter 3 Section 3.3.1, the knockout of *cwp84* at multiple sites resulted in a negative phenotype being obtained, regardless of the intron insertion location. Therefore, the methodology used in this chapter was to only progress one intron insertion site in each gene. As a result, not every chosen target was knocked out (Table 5.1.2). It is possible that knockout failed for a number of reasons:

- Knockout may have failed for technical reasons but the insertion site may have been successful.
- The insertion site was not suitable for intron insertion.
- Inactivation of that particular gene resulted in a lethal phenotype (as is putatively generated by the KO of *slpA* (discussed in 5.4.1.2)).

5.3.2 Genetic characterisation

PCR, using primers flanking the chosen genes, was used to confirm intron insertion in the Erm resistant *C. difficile* clone(s) (Figure 5.3.1).

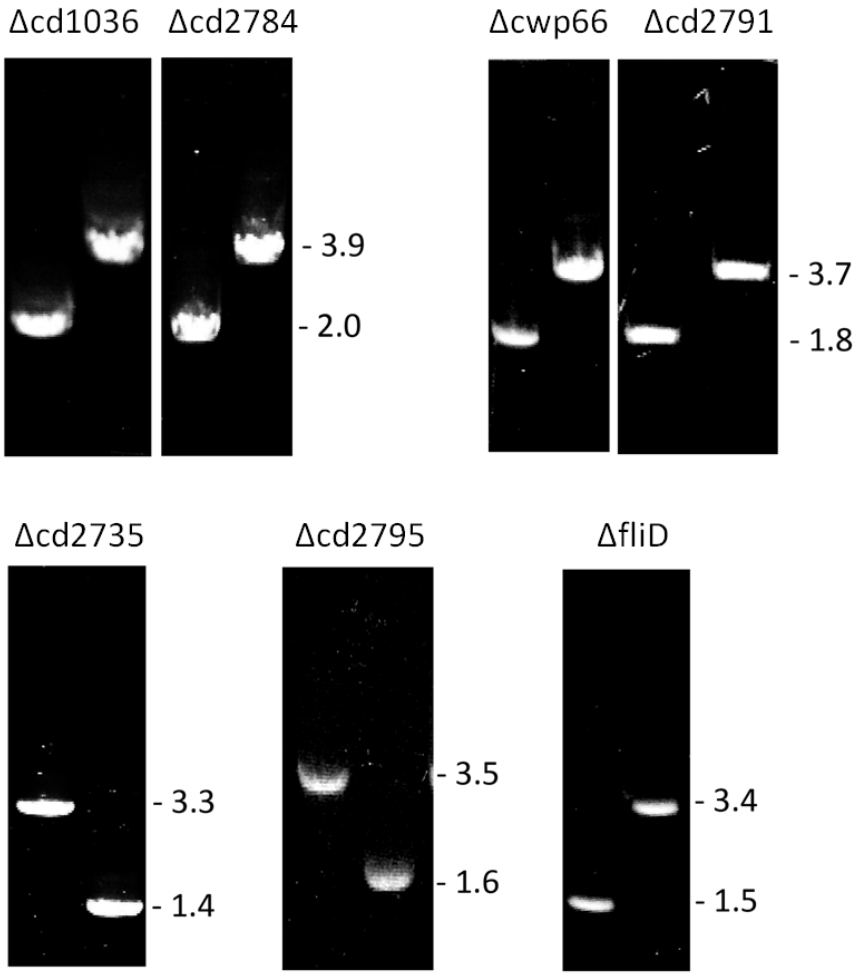


Figure 5.3.1 PCR confirmation of intron integration into selected surface protein genes. Lower band represents PCR product of WT, upper band represents amplification post intron insertion i.e. KO. Numbers refer to kilo base pairs.

5.3.2.1 Transcriptional analysis

Expression of *cd2791*, *cwp66*, *cd2784* and *fliD* have been described previously (Calabi *et al.* 2001; Tasteyre *et al.* 2001b; Savariau-Lacomme *et al.* 2003). To investigate the expression of *cd2795*, *cd1036*, *cd2735*, during normal growth, RT-PCR was performed on *C. difficile* 630ΔErm stationary phase RNA (Figure 5.3.2).

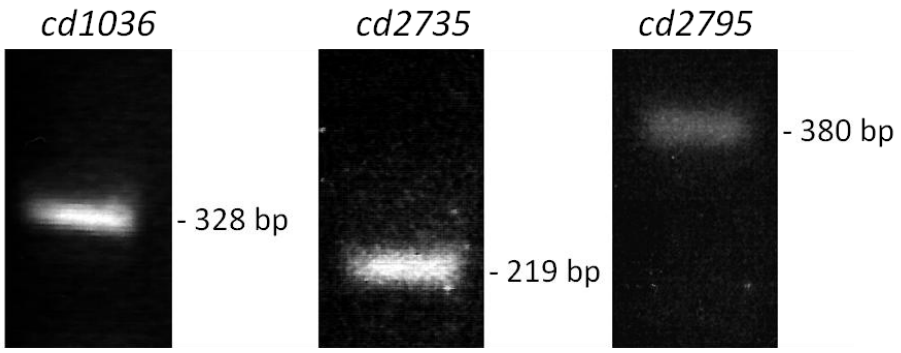


Figure 5.3.2 RT-PCR of *cd1036*, *cd2735* and *cd2795* from stationary phase *C. difficile* 630ΔErm RNA. The size of the reaction products reflects the regions chosen for amplification within each sequence.

Data shows that *cd1036*, *cd2735* and *cd2795* are all expressed in stationary phase in *C. difficile* 630ΔErm; amplification of *cd2795* gave a weak transcript using two different primer sets (Table 5.2.3). *In silico* analysis upstream or downstream of *cd1036*, *cd2735* and *cd2795* in the *C. difficile* 630 genome, suggests there are no genes within a short intergenic distance (<100 bp), therefore expression of *cd1036*, *cd2735* and *cd2795* is assumed to be monocistronic.

It is suggested that *cd2791-cd2790-cwp66* (*cd2789*) are encoded on a polycistronic message (Savariau-Lacomme *et al.* 2003). To confirm this, and to investigate expression this operon after intron insertion, RT-PCR was performed. Firstly, using primers flanking the intron insertion sites (Figure 5.3.4, primer set 1(a) or (b)), no products were obtained for the corresponding mutant (Figure 5.3.3 (i) and (ii)), confirming truncation of mRNA upon intron insertion

To understand how intron insertion, affected the polycistronic message, particularly in CDΔcd2791$_{1569}$, RT-PCR was performed amplifying the intergenic regions between *cd2791-cd2790-cwp66* (*cd2789*), similar to RT-PCR performed by Savariau-Lacomme *et al.* (2003) (Figure 5.3.3 (iii – v)).

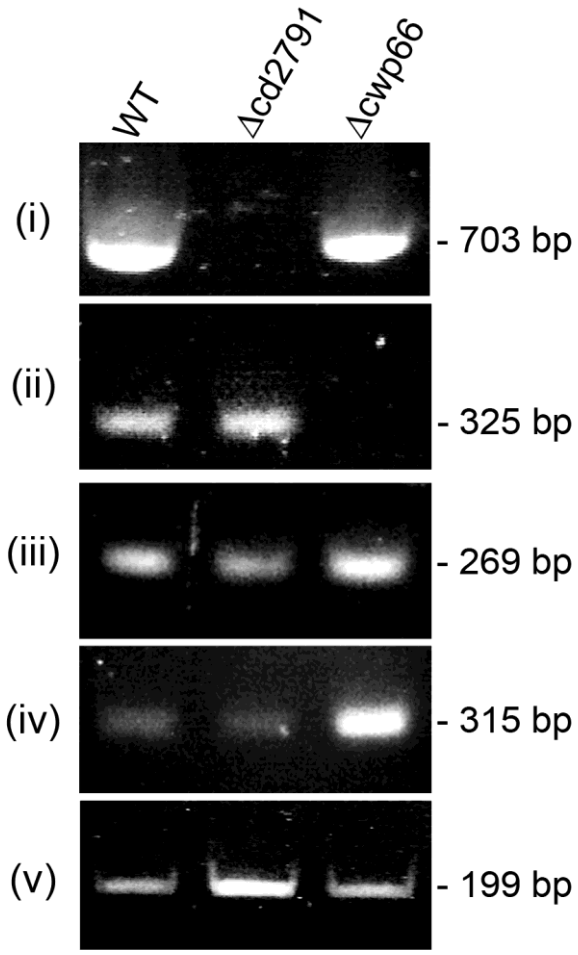


Figure 5.3.3 RT-PCR using stationary phase RNA from *C. difficile* 630ΔErm, CDΔ*cd2791*$_{1569}$ and CDΔ*cwp66*$_{519}$. Using: (i & ii) Primers amplifying *cd2791* or *cwp66*, flanking the intron insertion site (Figure 5.3.4 primers 1(a) or (b)). (iii) Primers amplifying the intergenic region between *cd2791-cd2790* (Figure 5.3.4 primers 2(a)). (iv) Primers amplifying the intergenic region between *cd2790-cwp66* (*cd2789*) (Figure 5.3.4 primers 2(b)). (v) Primers amplifying directly after the intron in CDΔcd2791$_{1569}$.The size of the reaction products reflect the regions chosen for amplification within each sequence.

Using intergenic primers (Figure 5.3.4, primer set 2(a) or (b)) a transcript is found, as expected for a polycistronic message, in *C. difficile* 630ΔErm and CDΔcwp66$_{519}$,

but also in CDΔcd2791$_{1569}$. A transcript is also found directly after the intron in CDΔcd2791$_{1569}$ (Figure 5.3.4, primer set 3). The presence of a post-*cd2791* transcript in CDΔ*cd2791*$_{1569}$ RNA, suggests a transcript, transcribed from within the intron (putatively from the *thl* promoter of the *ermB* gene), continues through the intergenic regions and putatively encodes a message for CD2790 and Cwp66.

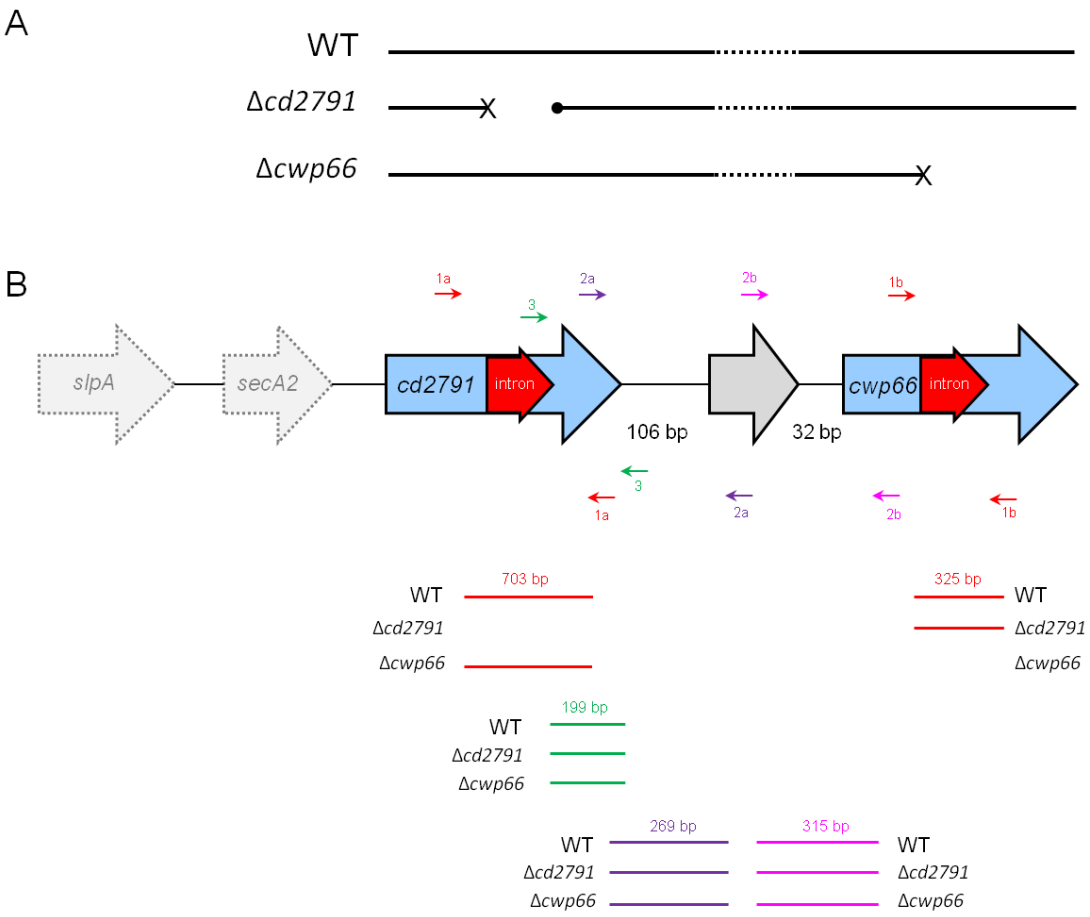


Figure 5.3.4 Schematic of RT-PCR analyses of *cd2791-cd2790-cwp66* cluster. (A) Proposed polycistronic message of *C. difficile* 630ΔErm, CDΔcd2791 and CDΔcwp66. X – termination of transcript. • – start of transcript, hashed line - assumed continuation of mRNA (although no RT-PCR performed). (B) Location of primers and size of expected products used in transcriptional analysis (not to scale). Primers are colour coded with both their respective pair and their product. Primer sets - 1(a) & (b) - flanking intron insertion sites, 2 (a) & (b) - intergenic primers (same as used in Savariau-Lacomme *et al*. (2003)), 3 - *cd2791* post intron insertion site.

5.3.3 Characterisation of the knockout mutants

5.3.3.1 Morphology and growth characteristics

5.3.3.1.1 Growth

All surface protein KOs appeared to have normal colony morphology and appeared similar to WT cells under Gram stain. The growth rate of the mutants in liquid medium was assessed (Figure 5.3.5). All mutants grew at a similar rate to the WT, except CDΔcd2784$_{438}$ (p=0.027, Paired T-test) which appeared to grow significantly slower than the WT. The flagella cap (fliD) KO while grew slower, was not statistically significantly different to the WT growth rate (p=0.053, Paired T-test).

5.3.3.1.2 Motility

The CDΔfliD$_{120}$ mutant had a tendency to settle in unstirred liquid culture vessels, suggesting an inability to maintain in suspension. Results show CDΔfliD$_{120}$ did not appear to demonstrate any motility into soft agar compared to the parental strain *C. difficile* ΔErm or control 001 ribotype strains, suggesting that CDΔfliD$_{120}$ is non-motile (Figure 5.3.6). This result is consistent with recently reported data, suggesting both the FliD and FliC mutants are non-motile (Baban *et al.* 2010).

5.3.3.1.3 Sporulation

Mutants appeared to retain their ability to sporulate with none of those assessed being significantly different to the WT (p>0.05, Students T-Test) (Figure 5.3.7) suggesting that the ability to form viable, heat resistant, spores was unaffected.

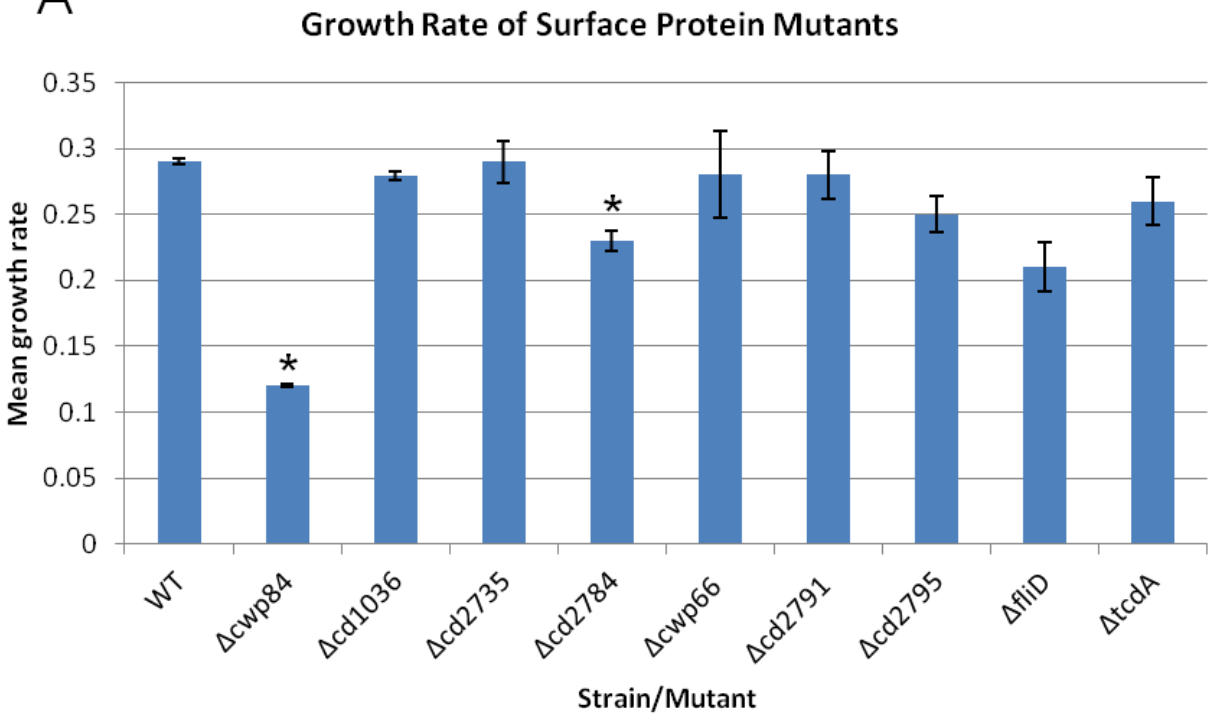
5.3.3.2 Toxin production

Toxins A and B are produced when cultures reach stationary phase (usually after 15 hr, 24 hr for *C. difficile* 630 (Merrigan *et al.* 2010)) and when cultures are stressed. The ability of surface protein KO mutants to produce toxin was assessed by ELISA and compared to toxin production of the wild-type under similar growth conditions.

Figure 5.3.8 shows 24 hr culture supernatants from surface protein mutants had higher toxin A concentrations than the wildtype (p>0.05, Students T-test) except CDΔcd2784$_{438}$ (p=0.59, Students T-test) as measured by ELISA. This increase was

particularly high in CDΔcwp84$_{347}$ and CDΔfliD$_{120}$. CDΔtcdR$_{499}$ did not appear to produce any toxin A above the limit of quantification (5 ng/ml), consistent with its role as a positive toxin regulator. Disruption of a gene key for sporulation, but not putatively surface located (*spo0A*), also results in significantly increased toxin production after 24 hr ($p<0.05$, Students T-Test), a result which is contrary to a previously reported *spo0A* KO (Underwood *et al.* 2009). Toxin B levels of mutant 24 hr culture supernatants, although demonstrated a slight increase, were not significantly different ($p>0.05$, Students T-test) to the wildtype.

A



B

Strain/Mutant	% WT
WT	100%
Δcwp84	42%
Δcd1036	96%
Δcd2735	99%
Δcd2784	80%
Δcwp66	96%
Δcd2791	95%
Δcd2795	86%
ΔfliD	71%
ΔtcdA	88%

Figure 5.3.5 Growth Rates of *C. difficile* surface protein mutants. (A) Growth rates measured on the linear portion of the growth curve (* denotes statistically significant ($p<0.05$) slower growth rate than the WT) (n=3) (B) Comparison of growth rate of mutants with WT as a percentage.

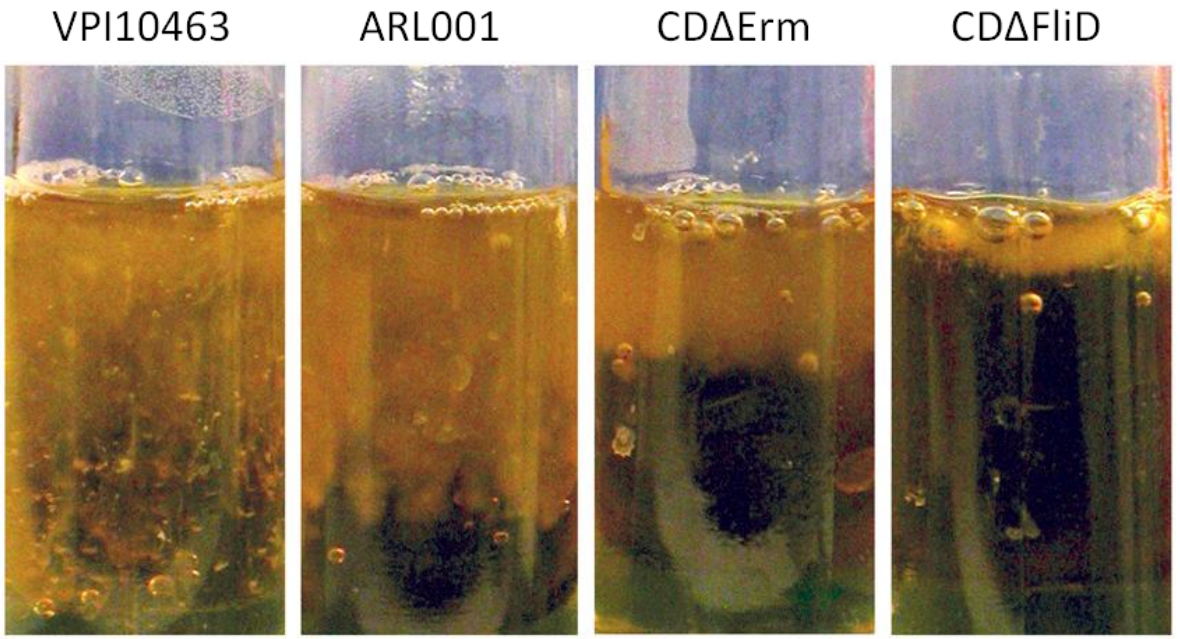


Figure 5.3.6 Motility of *C. difficile* strains in molten agar. *C. difficile* was inoculated into the first 5 mm of 0.175% sBHI agar and grown for 24 hr.

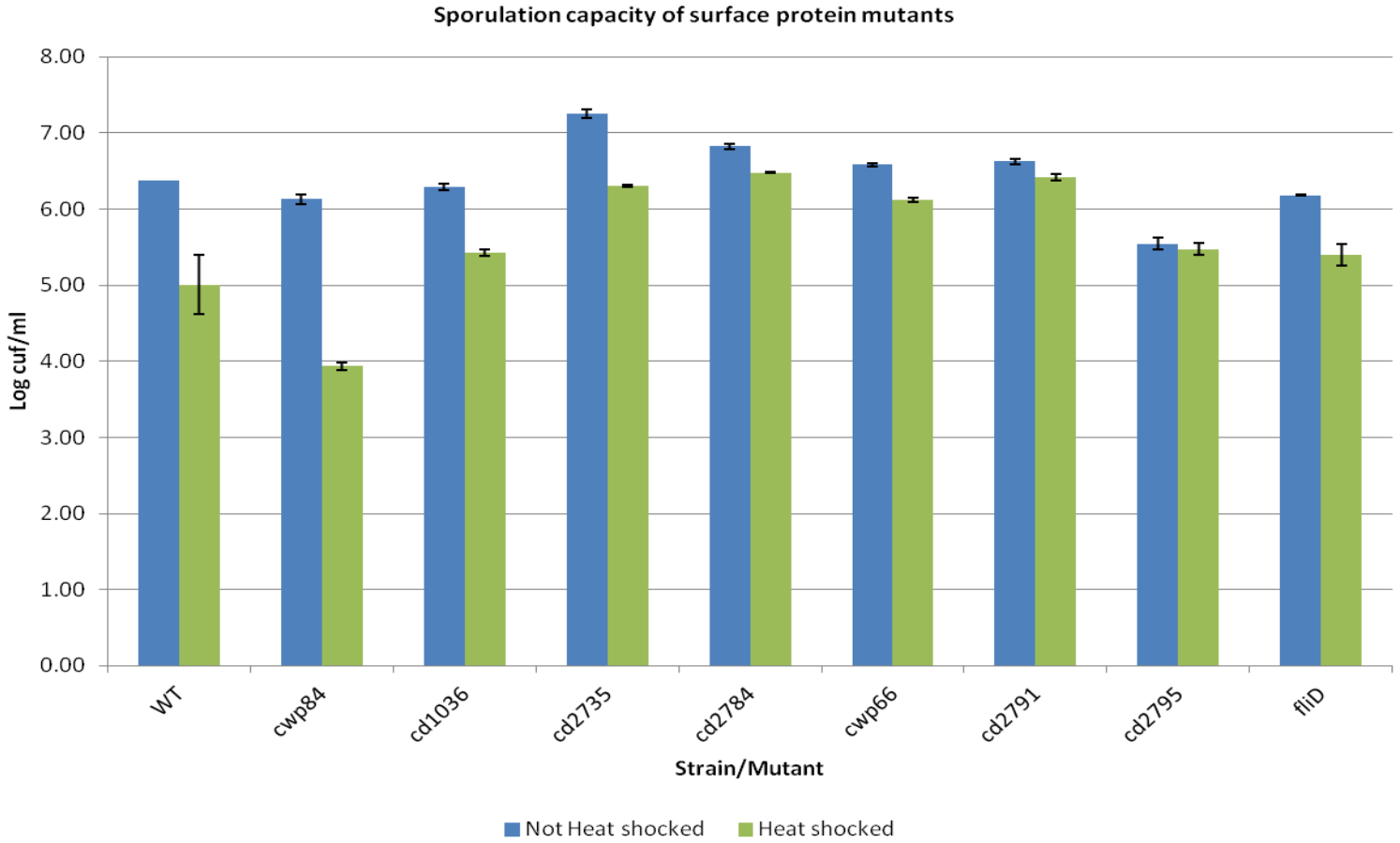


Figure 5.3.7 Viable spore count of surface protein mutants.

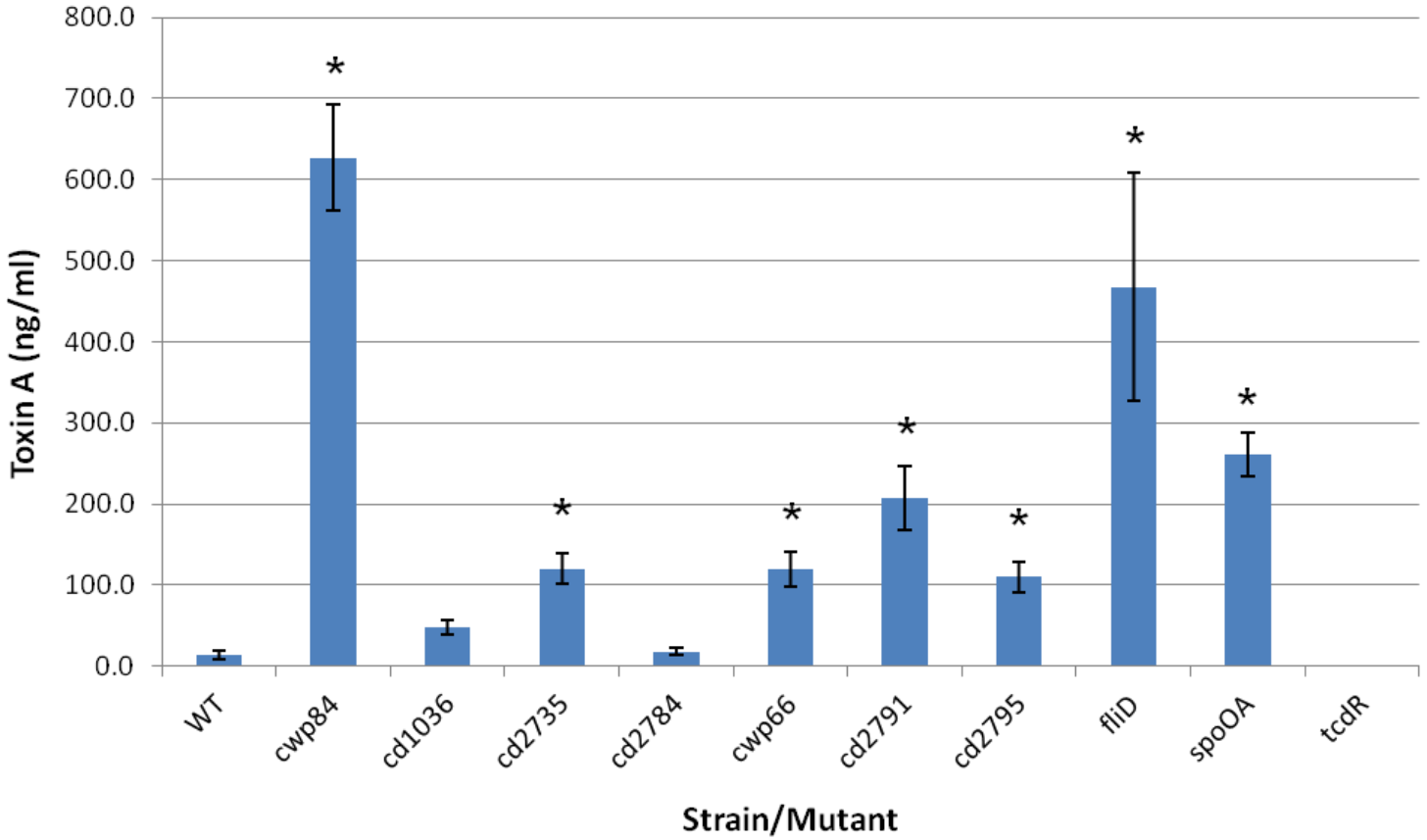


Figure 5.3.8 Estimation of Toxin A concentration in culture supernatant from *C. difficile* strains/mutants by ELISA. Strains grown in sBHI culture medium for 24 hr (* denotes statistically significant ($p < 0.05$) higher toxin A concentration than the WT). Limit of quantification 5 ng/ml.

5.3.3.3 Analysis of surface layer proteins of surface protein mutants

In order to determine if the knockout of various surface proteins had an effect on the surface of the bacterium, the surface proteins were removed by low pH extraction and examined by SDS-PAGE (Figure 5.3.9).

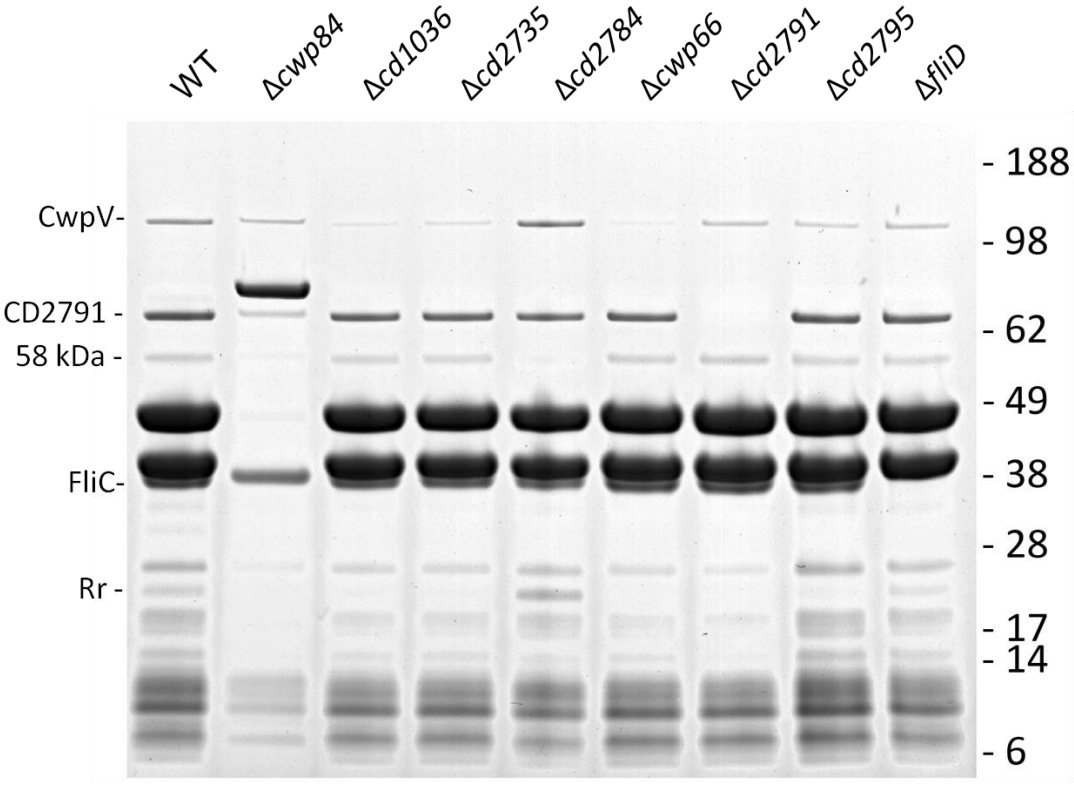


Figure 5.3.9 Coommasie stained SDS-PAGE gel of low pH SLP extracts from a range of surface protein mutants. FliC- Flagella core protein, Rr- rubrerythrin (Chapter 4 Section 4.3.1).

Mutant SLPs appear largely similar to the WT SLPs. The main S-layer bands appear unaffected and are present at 40 kDa and 47 kDa respectively. The lower MW fragmentation pattern (<28 kDa) is similar to the WT.

The SLPs of CDΔ*cd2791*$_{1569}$ lack a 66 kDa species, mass spectrometric analysis of this protein in the WT identified it as CD2791 (Cwp2), confirming the loss of the CD2791 from the cell surface corresponds with inactivation of *cd2791*. In CDΔcd2784$_{438}$ SLPs a 58 kDa band is less dense compared to other mutant and WT SLPs. The SLPs from

CDΔfliD$_{120}$ demonstrate a lack of a 38 kDa species (identified as FliC in CDΔcwp84 SLPs (Chapter 3 Section 3.3.4.1). The 115 kDa band (identified as CwpV in Chapter 3) appears unaffected in all mutant SLPs.

5.3.3.4 Adherence to the Caco-2 colonic cell line

Under standard culture conditions (after 14-21 days) Caco-2 cells spontaneously polarise and display presence both tight junctions, resulting in formation of functionally differentiated brush-border microvilli, thereby resembling enterocytes lining the small intestine. Caco-2 cells are widely used across the pharmaceutical industry as an *in vitro* model of the human small intestinal mucosa, to predict the absorption of orally administered drugs and used to assess adherence of gastrointestinal pathogens including *C. difficile* (Eveillard *et al.* 1993; Cerquetti *et al.* 2002). Therefore, using a cell based adhesion assay, the ability of the created surface protein mutants to adhere to Caco-2 cells *in vitro* was assessed (Figure 5.3.10).

Results show that mean adhesion of four mutants: CDΔ*cwp66*$_{519}$, CDΔ*cd2791*$_{1569}$, CDΔ*cd2795*$_{1320}$ and CDΔ*cd2735*$_{1113}$ was significantly lower than the WT ($p<0.05$, Students T-test). Two mutants CDΔcd2784$_{438}$ and CDΔfliD$_{120}$ displayed increased relative adhesion compared to WT. The data suggests that knockout of *cd2735*, *cd2795*, *cd2791* and *cwp66* decreases the ability of *C. difficile* to adhere to Caco-2 cells *in vitro*.

Due to the close association of *cd2791* and *cwp66* genetically, adhesion of CDΔ*cd2791*$_{1569}$ and CDΔ*cwp66*$_{519}$ to Caco-2 cells were assessed in a direct pair-wise comparison (Table 5.3.3). The mean adherence of CDΔ*cd2791*$_{1569}$ is statistically significantly less compared to CDΔ*cwp66*$_{519}$ ($p=0.003$, Students T-Test). Taken together, the data suggests that both Cwp66 and CD2791 play a role in adhesion, with the inactivation of *cd2791* having a profound effect on the ability of bacterial cells to adhere *in vitro* (Figure 5.3.11).

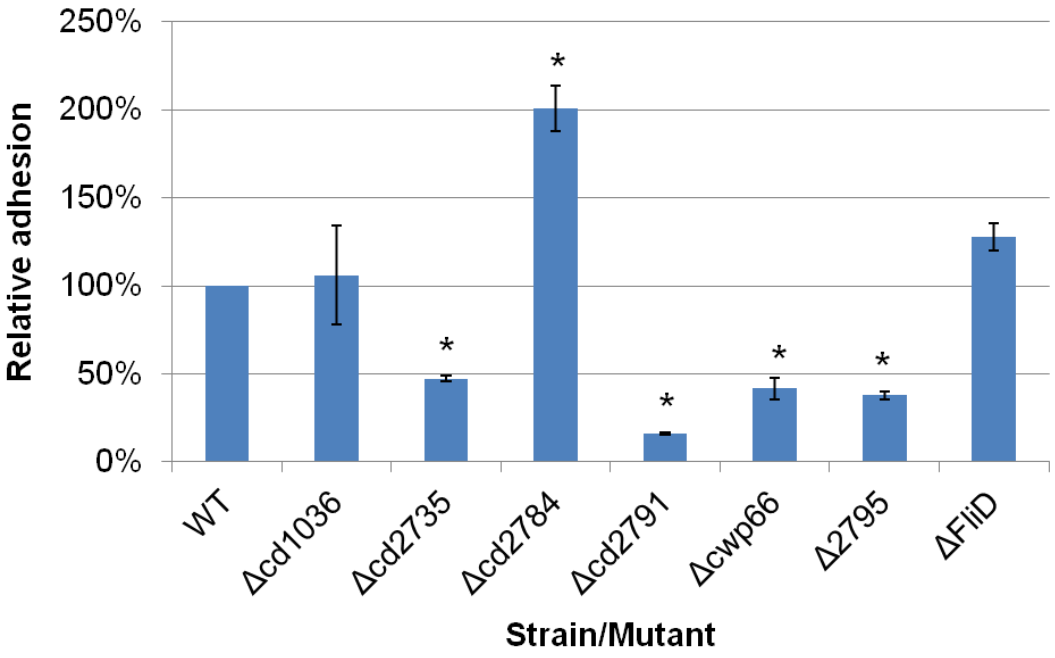


Figure 5.3.10 Relative adhesion of *C. difficile* surface protein mutants *in vitro*. Relative adhesion calculated from mean numbers of mutant *C. difficile* organisms adhering to a cell relative to *C. difficile* ΔErm in pair wise comparison (± standard error).

Measurements made by random field of view counts of a total of >150 Caco-2 cells vs bacterial cells at x1000 magnification, each adhesion assay was conducted in duplicate. * denotes statistically significant different adhesion compared to WT.

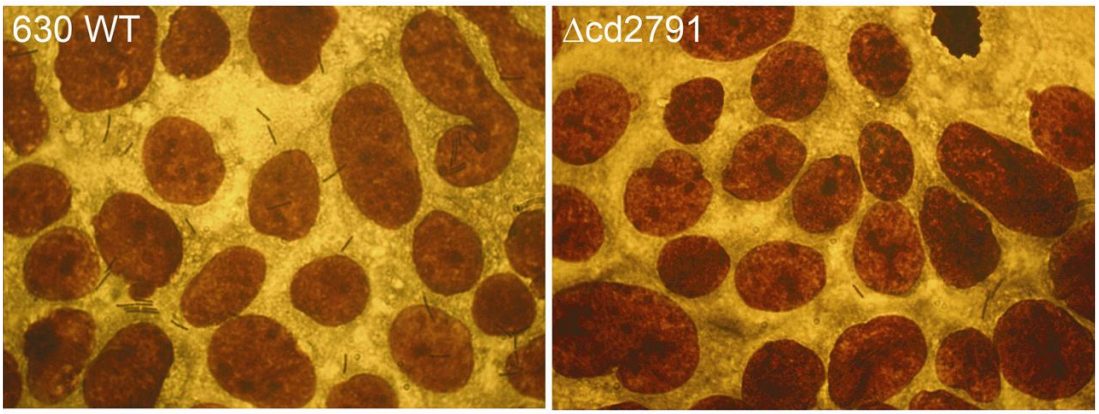


Figure 5.3.11 Random field of view comparision of adhesion of *C. difficile*ΔErm vs. CDΔ*cd2791*$_{1569}$ to Caco-2 cells in vitro. X1000 magnification.

Table 5.3.3 Adhesion of CDΔ*cd2791*$_{1569}$ vs. CDΔ*cwp66*$_{519}$ to Caco2 cells

Mutant	Adhesion*	% of Δ*cwp66* adhesion
Δ*cwp66*	0.15±0.03	100%
Δ*cd2791*	0.06±0.01	60.0±19.5%

*Mean numbers of *C. difficile* organisms adhering to a cell (± standard error). Measurements made by random field of view counts of a total of >150 Caco-2 cells vs bacterial cells at x1000 magnification, each adhesion assay was conducted in duplicate.

5.3.3.5 Binding of rCD2791$_{27\text{-}322}$ to cell lines

Since CDΔ*cd2791*$_{1569}$ displayed decreased adhesion to Caco-2 cells *in vitro*, the ability of a recombinant N-terminal domain of CD2791 (rCD2791$_{27\text{-}322}$) to bind to cell lines was assessed.

Unfortunately, the apparent auto-fluorescence of Caco2 cells prevented qualification of binding of Alexafluor488 labelled rCD2791$_{27\text{-}322}$. Using Vero cells binding of Alexafluor$_{488}$ labelled rCD2791$_{27\text{-}322}$ was observed at 1 µM and 50 µM (Figure 5.3.12). Unfortunately, the Vero cells were significantly more sensitive to the washing steps and despite duplicate repeats, Vero cells sloughed from 8-well chamber slides preventing confirmation of rCD2791$_{27\text{-}322}$ binding or competition assay.

On the basis of these preliminary results, the N-terminal domain of CD2791 (rCD2791$_{27\text{-}322}$) appears to have some adhesive properties *in vitro*.

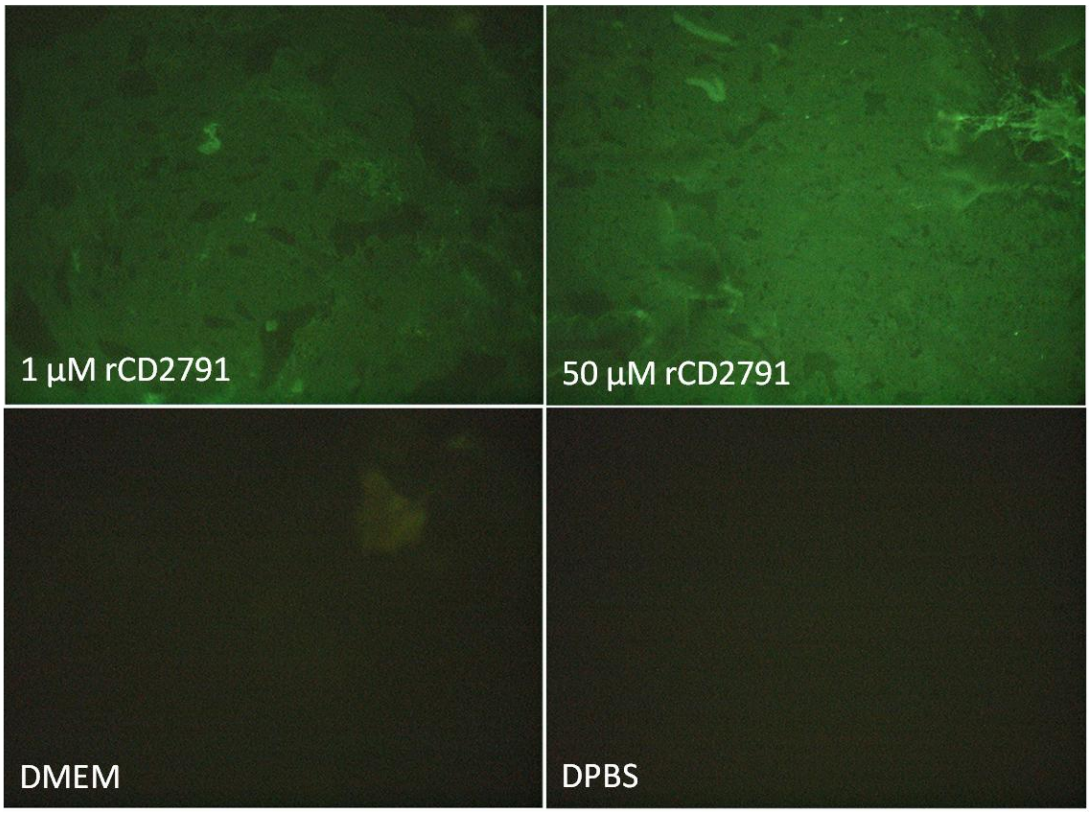


Figure 5.3.12 Adhesion 1 µM or 50 µM of Alexafluor$_{488}$ labelled rCD2791$_{27\text{-}322}$ to Vero cells. Slides viewed dry at x4000 magnification.

5.3.3.6 Purification of anti-CD2791 antibodies from anti-R20291 SLP IgG

As antibodies raised to SLPs react with a 66 kDa species (CD2791) in SLPs of any strain (Chapter 4 Section 4.3.3.1or 4.3.3.2), a method utilising an immobilised N-terminal fragment (rCD2791$_{27\text{-}322}$) to purify specific anti-CD2791 was assessed. These antibodies would then be used in the caco2 cell adhesion assay for their ability to prevent both rCD2791$_{27\text{-}322}$ fragment binding and *C. difficile* adhesion.

In order to quantify the level of the immune response to CD2791 in anti-R20291 SLP IgG, ELISA with immobilised rCD2791$_{27\text{-}322}$ was performed (Figure 5.3.13).

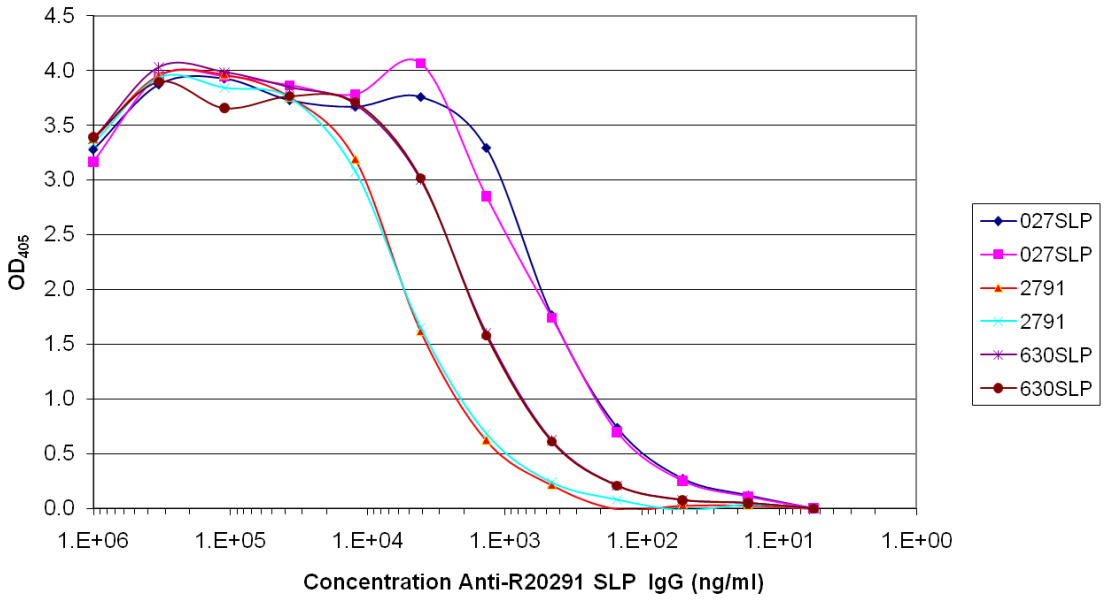


Figure 5.3.13 Binding of ovine anti-R20291 SLP IgG to rCD2791$_{27\text{-}322}$ in ELISA

Anti-R20291 SLP IgG reacted with 10 µg/ml rCD2791$_{27\text{-}322}$ in ELISA approximately 1 log less than corresponding SLPs and ½ log less than a differing strain SLPs (*C. difficile* 630ΔErm), suggesting that a significant proportion of the anti-SLP immune response is directed to the N-terminus of CD2791.

To purify specific anti-CD2791 antibodies from anti-R20291 SLP IgG, rCD2791$_{27\text{-}322}$ was coupled to N-hydroxysuccinimide (NHS)-agarose, with an approximate 60% coupling efficiency. Unfortunately, despite application of 40 mg of anti-R20291 SLP IgG (1 mg/ml), only 1 ml of 5 µg/ml specific CD2791 antibodies were eluted, too little for use in the adhesion assay.

5.4 Discussion

It is only relatively recently that a stable targeted gene knockout system for the genus Clostridium has been developed (Heap *et al.* 2007; Heap *et al.* 2010). The ClosTron system, derived from the TargeTron system (Sigma-Aldrich), inactivates a gene by targeted insertion of a group II intron (Ll.LtrB) from *Lactococcus lactis*. The Ll.LtrB intron is useful, as it has a broad host range (Gram positive and negative organisms) (Cousineau *et al.* 1998; Yao *et al.* 2005) and because the intron encoded protein (IEP) is expressed downstream of the Ll.ltrB intron (on a plasmid), the resulting insertion is non-splicing, i.e. is extremely stable. This has been demonstrated by Frazier *et al.* (2003), who found no loss of bacteriophage C2 resistance of an M1083s-opt::abiD integrant after 80 passages.

To date, there are only four publications reporting the knockout of surface associated proteins (CwpV (Emerson *et al.* 2009a), Cwp84 (Kirby *et al.* 2009) , FliC and FilD (Baban *et al.* 2010) and Fbp68 (Lin *et al.* 2010)). There is therefore considerable scope for furthering knowledge of *C. difficile* pathogenesis via a gene KO approach. Particularly for surface associated genes which possess a CWBD which putatively attaches them (non-covalently) to the cell surface. In this chapter a selection of surface associated genes (PFam 04122 containing) and the flagella cap were knocked out and the mutant strains characterised.

5.4.1 Genetic considerations of *C. difficile* knockouts

5.4.1.1 Production of *C. difficile* knockouts

A total sixteen genes (22 insertion sites) were selected for knockout (Table 5.1.2), eleven contained an N- or C- terminal CWBD, two were surface associated (FliD and CD2797) and three were putative pseudogenes (annotated as such in the *C. difficile* genome), to assess whether any phenotypes observed were not directly related to the ClosTron protocol or intron integration.

The algorithm for selecting insertion sequences and retargeting the region responsible for intron specificity was initially only available via the Sigma TargeTron

website, but subsequently became available free of charge via www.clostron.com (Heap *et al.* 2010). As revealed in Table 5.3.2, intron insertion sites given by the Sigma design site and the those given by the ClosTron.com (Perutka *et al.* 2004) algorithm are different. Out of the 12 sites (8 genes) which were successfully knocked out, 8 sites were retargeted via TargeTron and 4 sites were retargeted by ClosTron .com (Table 5.1.2). On the basis of these results, KO success is more likely with the Sigma TargeTron algorithm than with the ClosTron.com algorithm.

Out of all the sixteen genes selected for knocked, eight were not able to be knocked out; these were *slpA*, *cd1751, cd2713, cd2767, cd2797* and three pseudogenes. There are several reasons to explain failure to KO selected genes in this chapter.

The first of these, is the approach taken in this chapter (as opposed to that used in Chapter 3), whereby choosing only one target site in each gene severely restricts the likelihood of success. If a chosen target site is not amenable to intron insertion then the gene will not be knocked out, however many times the site is tried, i.e. independent of protocol repetition or optimisation. The second reason for KO failure may be technical problems with specific parts of the ClosTron protocol, e.g. conjugation, or reversion of *C. difficile* 630ΔErm to erythromycin resistant (as reported by (Hussain *et al.* 2005)), making integrant screening difficult. A third reason for failure may simply be the screening of an insufficient number of clones for correct integration. Heap *et al.* (2009a) recommends that if screening 4 clones does not result in successful integrant identification, to flood the plate and/or screen large amounts of colonies. The third reason is not failure to KO, but failure to isolate KO clones. Finally, and particularly applicable to *slpA* (discussed in 5.4.1.2 below) is the possibility that insertional inactivation of the chosen genes results in a lethal phenotype.

A trial of a double KO using CDΔ*cd2791*$_{1569}$ and CDΔ*cwp66*$_{519}$ or vice versa was unsuccessful, however, to date, only one double KO mutant has been reported (Kuehne *et al.* 2010) and utilised a slightly modified knockout procedure. It is also

possible that the double *cd2791/cwp66* mutant failed due to alterations on the cell surface prevented successful conjugation.

5.4.1.2 Knockout of *slpA*

Knockout of the main S-layer gene was repeatedly attempted; however, no mutant clones were isolated. It is suspected that KO of *slpA* creates a lethal phenotype, in which the cell can no longer retain any structural integrity, as provided by a rudimentary S-layer (such as that found in the *cwp84* mutant (Kirby *et al.* 2009)). However, S-layer negative strains of *Aeromonas sp*. (Janda *et al.* 1994), *B. anthracis* (Kern & Schneewind 2009), *Tannerella forsythensis* (Sakakibara *et al.* 2007) or *Campylobacter fetus subsp. Fetus* (Grogono-Thomas *et al.* 2000) exist, suggesting that an S-layer negative *C. difficile* mutant may be possible, as normally S-layer positive bacteria can survive without their S-layers.

As the KO of certain surface protein KOs resulted in earlier release of toxin A and B, it was hypothesised that *tcdE* (a holin of *C. difficile* that lyses *E. coli* (Tan *et al.* 2001)) may also be upregulated. Therefore, a double *slpA*:*tcdD(R)* knockout was tried, *tcdD(R)* is a positive (alternative sigma factor) regulator of toxin production (Karlsson *et al.* 2003). By using a *tcdD(R)* mutant, putative upregulation *tcdE* or other self harming proteins under *tcdD(R)*'s regulation should not be induced. This technique created pin prick transconjugants previously unattainable using the normal *C. difficile* 630ΔErm; however, subsequent plating dramatically decreased colonies, indicating that while conjugation could occur upon intron integration a lethal phenotype resulted. The type of colonies obtained may be similar to those obtained by knockout of glycerol dehydrogenase (*gldA*) by Liyanage *et al.* (2001). Dang *et al.* (2010) also highlights how, to date, despite widespread efforts it has not proven possible to generate a *slpA* knockout strain.

Preventing the lethal phenotype produced from KO of essential genes is therefore a challenge. Creation of conditional mutants or complementation prior to KO may provide a way to circumvent the effects of inactivation. If the latter technique is

used, particularly with the ClosTron technique, one must prevent the intron inserting into the plasmid rather than the desired location in the host genome.

5.4.1.3 Transcriptional analysis of *cd2791-cd2790-cwp66* cluster

RT-PCR in this chapter has confirmed the existence of a polycistronic transcript between *cd2791-cd2790-cwp66*, as suggested by Savariau-Lacomme *et al.* (2003). Transcription of the polycistronic message is assumed to proceed through *cd2791-cd2790-cwp66* in *C. difficile* 630ΔErm and up until the intron in CDΔcwp66$_{519}$ (Figure 5.4.1, WT and Δ*cwp66*). Due to the polar insertion of the intron in CDΔ*cd2791*$_{1569}$, a transcript, originating from the *thlA* promoter (used to drive the introns ErmB expression), is found past the intron insertion site putatively encoding CD2790 and Cwp66 (Figure 5.4.1, Δ*cd2791*).

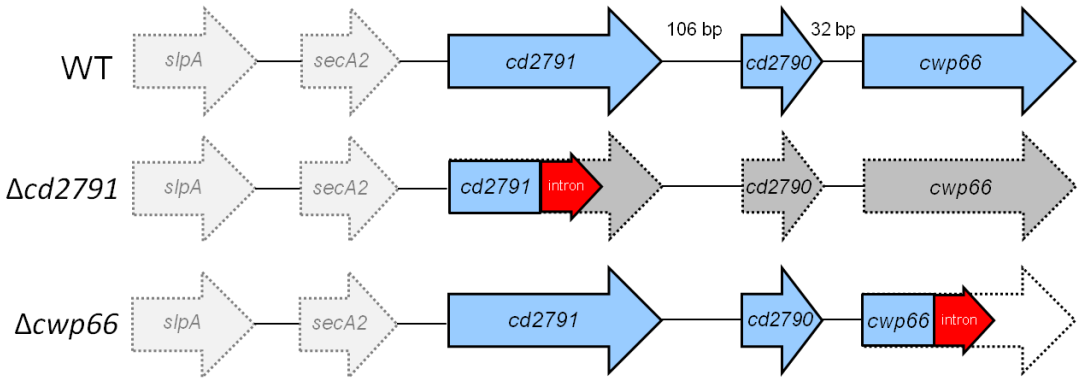


Figure 5.4.1 Proposed transcription of the *cd2791-cd2790-cwp66* operon before and after knockout of *cd2791* and *cwp66*. All three genes are encoded as a polycistronic message (light blue) in the wildtype *C. difficile* 630ΔErm (WT). Messenger RNA is truncated at *cwp66* in CDΔcwp66$_{519}$ (Δcwp66) and *cd2791* in CDΔ*cd2791*$_{1569}$ (Δcd2791); however in CDΔ*cd2791*$_{1569}$, an intron derived transcript continues along the operon and putatively encoding a message for CD2790 and Cwp66 (grey).

It is likely that unless CD2790 and Cwp66 are post-transcriptionally controlled, polar intron insertion in CDΔ*cd2791*$_{1569}$ results in non-wildtype transcription of *cd2790* and *cwp66* and may subsequently alter CD2790 and Cwp66 levels. This polar effect

could be resolved by removal of the *ermB* gene using FRT sites, leaving only the 940 bp group II intron region, which is normally efficiently transcribed and should not cause premature termination of transcripts (Heap *et al.* 2009a), or by using a unique gene deletion technique recently developed using the ClosTron system (Jia *et al.* 2011).

It is unknown as to whether the genes downstream of *cd2791* are still translated from the resultant (intron-derived) message. However, it is likely that the regulation of *2790-cwp66* and putatively the resulting protein(s) is altered from the WT (unless they are post-transcriptionally regulated). The identification/presence of Cwp66 on the cell surface of CDΔ*cd2791*$_{1569}$ and/or proof that its regulation is unaffected by *cd2791* KO would be required to confirm phenotypic changes were not due to decreased expression of *cwp66* (or *cd2790*). Although, inactivation of *cd2791* decreases adhesion more than *cwp66* KO alone. This suggests that CD2791 plays a role in adhesion different to that of Cwp66, or CD2790 plays a larger role than previously thought in mediating adhesion.

CD2790 is labelled a hypothetical protein in the *C. difficile* 630 genome but shares amino acid homology with N-acetylglucosaminyl phosphatidylinositol deacetylases, which participate in the second step of glycosylphosphatidylinositol (GPI) biosynthesis. GPI is used to anchor surface proteins in membranes in eukaryotic cells (Kinoshita & Inoue 2000). However, HMM-HMM structure modelling reveals that CD2790 shares strong homology to the zinc-dependent metalloenzyme deacetylases. Deactelyation of peptidoglycan protects the bacterial cell wall from host lysozymes (Blair *et al.* 2005; Wang *et al.* 2010). The role of a deacteylase in conjunction with CD2791 or Cwp66 in *C. difficile* is currently unknown, but may be connected with a surface based response to stress.

Clusters of genes modifying cell surface proteins have been found in *Shigella flexneria*, where a cluster of three genes (*gtrA,B,X*), in a series of steps, glucosylate the O antigen (Guan *et al.* 1999). Hence, given the polycistronic message, it is possible *cd2791-cd2790-cwp66* form a part of the bacterial surface operon.

However, as discussed in Chapter 4, as *cd2791-cd2790-cwp66* are encoded on the same transcript, the lack of identification of Cwp66 during normal growth (Wright *et al.* 2005) and the constitutive presence of CD2791 on the cell surface, suggests these two proteins are differently (possibly post-transcriptionally) regulated.

5.4.2 Characterisation of the knockout mutants

5.4.2.1 Colony morphology, growth, motility and sporulation.

The colony morphology of all mutants produced in this chapter appeared WT-like, suggesting that inactivation of the chosen surface proteins do not affect the ability of the bacterium to form 'normal' spreading grey glass-like colonies. *C. difficile* colony morphology has been reported to change in response to Cwp84 or CwpV KO (Kirby *et al.* 2009; Reynolds *et al.* 2010). In CDΔCwp84, this is due to an increase in extracellular debris changing cellular migration and optical diffraction.
The growth rate of the mutants, were largely the same as the WT. The growth rate of the CD2784 mutant was significantly slower than the WT, suggesting that the ability of the bacterium to grow was affected.

It has been reported that inactivation of the peptidoglycan hydrolase, Acp, affects cell separation in *Clostridium perfringens* (Camiade *et al.* 2010). Therefore, it was expected that mutants deficient in CD2784 and CD1036, which contain a peptidoglycan aminohydrolase domain, would show altered cell morphology. However no difference was seen, suggesting that these proteins do not play a role in cell morphology or daughter cell separation.

As aforementioned *C. difficile,* in common with *E. coli,* has several peptidoglycan hydrolases. Therefore, inactivation of one such protein, e.g. CD2784, may not have an effect, as another can fulfil its role e.g. CD1036. In 1976, Fein & Rogers (1976) reported *B. subtilis* N-acetylmuramoyl-L-alanine amidase and endo-beta-N-acetylglucosaminidase mutants appeared to grow at normal rates, but as very long chains of un-separated cells. The mutants readily reverted, suggesting that the mutation was in a regulatory gene rather than the autolytic enzymes themselves. Therefore, it is possible that inactivation of an, as yet undiscovered, autolytic

regulatory protein(s) will have more of a marked effect on *C. difficile* than individual peptidoglycan hydrolase mutants. Work by Johann Peltier and colleagues is investigating this hypothesis by inactivation of *sigD*, a putative regulator of autolysins (Peltier *et al.* 2010). Indeed, all the twelve *E. coli* periplasmic peptidoglycan hydrolases have the potential to degrade the cell wall, however why they do not under normal growth conditions still remains to be answered (Vollmer *et al.* 2008).

All mutants produced in this chapter appeared to sporulate with sporulation level similar to that of the WT. Cell wall hydrolase mutants of *Streptomyces coelicolor* have heat sensitive spores (Haiser *et al.* 2009), however, all mutants produced heat resistant spores in this chapter.

Consistent with both the FliD KO in *Pseudomonas aeruginosa* (Arora *et al.* 2000) and the FliD mutant produced by Baban *et al.* (2010), the FliD mutant produced in this study was non-motile. The non-motile nature of the flagella cap (FliD) mutant caused it to appear to grow slower in un-agitated liquid cultures (as measured by A_{600}), as cells settled rather than producing normal suspension.

5.4.2.2 Toxin production of mutants

Toxin production in *C. difficile* is initiated in times of nutritional stress, typically occurring in stationary phase (Karlsson *et al.* 2008; Merrigan *et al.* 2010). Toxin production also appears to be increased upon exposure to sub-MIC concentrations of certain antibiotics (Onderdonk *et al.* 1979; Gerber *et al.* 2008). However, to date, there is only one reported occurrence of increased toxin production after knockout of surface associated genes (Baban *et al.* 2010).

Culture supernatants of CD2735, Cwp66, CD2791, Cwp84, CD2795, FliD and Spo0A mutants contained more toxin A than the wildtype, *C. difficile* 630ΔErm, after 24 hr. As discussed in Chapter 3, the increased toxin production found for the *cwp84* mutant is likely to be a result of the altered S-layer and the upregulation of stress pathways due to decreased cell integrity, or that decreased cell integrity means

toxin is able to 'leak' from the cell (a mechanism similar to the role putatively performed by TcdE (Tan *et al.* 2001))

As CD2791 is surface located, it is possible that removal of CD2791 from the cell envelope results in decreased cell integrity or results in changes to membrane/cell wall genetic regulation of toxin genes, resulting in earlier release of toxin. The current lack of identification of CD2735, CD2795 on the cell surface and the conditional expression of Cwp66 draws the cell integrity hypothesis into question. One presumes that the cell wall of the *spo0A* mutant is not affected, yet the *spo0A* mutant in this chapter produces increased toxin similar to CD2735, Cwp66, CD2791 and CD2795 mutants. Although, caution is to be exercised regarding *C. difficile* Spo0A mutants and toxin production, as there is conflicting evidence regarding links between sporulation and toxin production, as results in this chapter contradict Underwood *et al.* (2009). A *Bacillus subtilis spo0A* mutant was prone to cell lysis (though a lack of autolysin degradation via extracellular proteases) (Kodama *et al.* 2007), therefore, although Spo0A is not directly surface based, it may still play a role in cell integrity and as a result, on toxin production.

The non-motile nature of the FliD KO (or the FliC KO in (Baban *et al.* 2010)) means the bacterium is limited to utilising the nutrients in the local area, once these are spent, i.e. nutrient limitation conditions, toxin production begins. Baban *et al.* (2010) also suggests that the increased toxin may help the bacterium cause disease in the animal model, as postulated for the Cwp84 KO.

Together these results suggest that certain surface proteins have an intimate link with toxin production, exemplified by the CD2784 mutant not producing any more toxin A than the WT after 24 hr. Therefore, it appears CD2784 is not involved in any pathways which result in the release of toxin, particularly if they become non-functional.

It is interesting that *C. difficile* 630 (and its ΔErm variant) are low toxin producers, whereas ribotype 027/B1 strains have increased toxin production, as does VPI10463

(Merrigan *et al.* 2010). Do these high toxin producing strains have an altered surface that results in increased toxin production, increasing their virulence? Are toxin production, cell metabolism and cell surface integrity connected via a regulatory network e.g. a change in CodY expression (Dineen *et al.* 2010)? If so, the integration of these two systems may be key to infection progression *in vivo*. Moreover, co-administration of crude *C. difficile* toxin preparations, with a non-toxigenic strain, resulted in a significantly greater adherence to small bowel mucosa comparable to that of the toxigenic strain, suggesting the ability to produce toxin may confer some ecological, colonisation, advantage (Borriello & Barclay 1985; Borriello *et al.* 1988; Feltis *et al.* 1999).

5.4.2.3 Surface proteins of mutants

The SLPs of mutants produced in this chapter are largely similar to the SLPs of the WT. The differences that do occur are the loss of the 66 kDa CD2791 protein (CD2791 KO), loss of the 36 kDa FliC (FliD KO) and changes in a 58 kDa band intensity (CD2784 KO). The loss of the FliC band from the SLPs of FliD indicates that the flagella are no longer correctly assembled, confirmed by a lack of motility in the FliD KO.

Confirmation of KO i.e. the loss of a knocked out protein from the cell surface or identifying if the original KO affects other surface proteins, is largely dependent on the use of a method (growth and extraction) which reliably removes that protein(s) from the WT. Therefore, particularly applicable to CD2795 and CD2735 KO, confirming protein loss after KO is hampered by the apparent lack of these proteins on the cell surface. Therefore, cell surface based changes may have occurred in the mutants produced in this study, but using low pH glycine extraction they have not been revealed. The lack of observed major effect on the SLPs upon loss of Cwp66, CD2795, CD2735 and CD1036, could suggest how their presence and expression is conditional.

Antibodies to the surface proteins knocked out in this chapter may have helped their identification (and their subsequent loss). Interestingly, Abcam plc (Cambridge

UK) sell two antibodies (catalogue numbers 93756 and 93728), one to the N-terminal and another to the C-terminal end of a protein “highly similar to 1035”, if these are CD1036, they could have been used for immunological studies.

Confirming the loss of a knocked out protein (on the cell surface) is, however, only one part of the story. More complex methods such as microarray, proteomics or transcriptome sequencing are likely to be required to reveal, or start to reveal, the complex interactions between surface proteins and the bacterium.

The ClosTron system is, therefore, best suited to where there is a prior idea of the likely outcome of KO or at least a way of measuring the KO outcome, rather than randomly targeting individual genes and hoping for a negative phenotype, as attempted in this chapter. A random mutagenesis approach, such as the use of a transposon (Cartman & Minton 2010; Hussain *et al.* 2010) may be more applicable to identification of genes required in key pathways.

5.4.2.4 Adherence to cell lines

Surface protein extraction techniques only reveal proteins found on the cell surface *in vitro* and do not demonstrate those which are only necessary for infection *in vivo*.

In the absence of *in vivo* animal models, an *in vitro* adhesion assay using Caco-2 cells was used to determine if any of the mutants produced in this chapter had altered cell binding abilities. It was discovered that using post confluent fully differentiated Caco-2 cells, four surface protein mutants (CD2735, CD2795, CD2791 and Cwp66) adhered lower than *C. difficile* 630ΔErm, indicating that inactivation of either one of these four proteins, results in a decreased ability to adhere to Caco-2 cells *in vitro*.

5.4.2.4.1 CD2735

CD2735 contains two SH3 domains which participate in protein-protein interactions, mediating assembly of specific protein complexes via binding to a proline-rich (PXXP) core-conserved binding motif in their respective binding partner

(Mayer 2001). The KO of CD2735 containing these domains and the subsequent decrease in adhesion suggests that the SH3 domains in CD2735 partially mediate binding in *C. difficile*. However, SH3 domains bind their ligand with low affinity, thus are enhanced by multiple interactions, suggesting that CD2735 is prevalent on the cell surface; however, it has not been found to date.

Putative roles for CD2735 include:

- Assisting in the formation of multi-meric protein complexes on the surface of *C. difficile*.
- Binding directly to host protein(s) sufficiently to anchor *C. difficile* to the host cell.
- CD2735 upon binding to a host protein(s) affects the host in such a way that adhesion is increased, but is not involved directly in adhering bacterial cells.
- CD2735 upon binding to a host protein affects *C. difficile* surface proteins such that other stronger adhesins (perhaps forming a complex) are brought into play.
- CD2735 is a signal molecule for *C. difficile* and its KO prevents a signal reaching an adhesin that would normally be upregulated in the presence of a binding environment.

It has been shown that invasive pathogens, such as *Salmonella typhimurium,* promote assembly of focal adhesion complexes, which contain SH3 domains, in response to bacterial attachment (Ireton & Cossart 1998; Shi & Casanova 2006). Moreover, modified SH3 domains are found on the surface of the *Listeria monocytogenes* invasion protein InlB and mediate binding to host ligands (Marino *et al.* 2002). While *C. difficile* is putatively not invasive, it is possible one or more of its surface proteins promote adhesion by modification of host cell physiology. The subversion of host cell physiology is found in the enteropathogenic *Escherichia coli* (EPEC) secreted protein, EspF. The EspF protein contains three proline-rich repeats and six putative SH3 binding domains and disrupts intestinal barrier function through mitochondrial targeting, specifically by binding to sorting nexin 9 (Marches

et al. 2006). Further work examining CD2735 and binding to host cell receptors is therefore required.

5.4.2.4.2 CD2795

The proposed ability of the bacterial Ig-like domains to interact (Potapov *et al.* 2004), suggests that loss of CD2795 from the cell surface may decrease adhesion through an inability to form CD2795-CD2795 and/or CD2795-CD2794 (adhesin) complexes, which may or may not directly impact interactions with host cell proteins. In Leptospira spp. three bacterial Ig-like domain containing proteins (LigA, B, C containing thirteen and twelve Ig-like domains respectively) have been found to bind fibronectin, and when administered to hamsters prevent mortality (Silva *et al.* 2007; Lin *et al.* 2008). However, given that these (Lig) proteins contain far more Ig-like domains than CD2795 (and CD2794), inferences about CD2795 (or CD2794) are limited, but serves to highlight how these domains may play a role in bacterial adhesion.

The role of the SCP domain in bacterial pathogenesis or adhesion is not well understood. The SCP family includes plant pathogenesis-related protein 1 (PR-1), mammalian cysteine-rich secretory proteins (CRISPs) and allergen 5 from vespid venom (Yeats *et al.* 2003). One protein, human GAPR-1 protein (PR-1 member), is suggested to form a serine protease active site upon dimerisation (Serrano *et al.* 2004). It is interesting that CD2795 is next to CD2794 which shares domain structure, if not particularly high amino acid identity (70.4%). It is possible CD2795 and CD2794 form a dimeric adhesin/protease complex key to binding to host cell receptors.

The apparent absence of CD2795 (or CD2794) on the cell surface, yet their loss from the cell decreases the ability of the bacterium to adhere *in vitro*, suggests how these and other proteins, not yet characterised, play a role in adhesion and possibly a role in adherence *in vivo*. Further research into both bacterial SCP domains and CD2795 (and CD2794) of *C. difficile* is therefore needed.

5.4.2.4.3 Cwp66

The Cwp66 protein has been proven to be an adhesin (Waligora *et al.* 2001). The results in this chapter confirm this, as knockout of this protein resulted in a decreased ability to bind to Caco2 cells. Waligora *et al.* (2001) suggests that Cwp66 is a heat shock inducible adhesin and found that anti-Cwp66 N- or C-terminal antibodies did not inhibit adherence if bacteria were not heat shocked. If Cwp66 was therefore a heat shock only inducible adhesin, knockout of Cwp66 should, theoretically, only affect the bacterium after heat shock and not under normal conditions. However, adhesion of the Cwp66 KO was less than 50% that of the WT under normal conditions, suggesting that Cwp66 plays a role in adhesion in normal growth conditions and that its expression is not associated with heat shock.

5.4.2.4.4 CD2791

The highest decrease in adhesion to Caco2 cell *in vitro* was that of the CD2791 mutant. The results obtained with the CD2791 mutant are difficult to interpret due to two factors. Firstly, the effects of knocking out a protein at the start of a polycistronic message (encoding *cd2791-cd2790-cwp66*). Secondly, applicable to CDΔ*cd2791*$_{1569}$, is presence of a powerful *thlA* promoter which can lead to increased expression of downstream genes or lead to increased expression of upstream genes via anti-sense and transcriptional interference effects (Heap *et al.* 2009a) i.e. non-wildtype levels of protein or message, due to non-wildtype regulation. As discussed in Section 5.3.2.1, a solution to this problem is the removal of the *ermB* gene, flanked by FRT sites, leaving only the 940 bp group II intron which putatively has no effect on downstream genes. This technique requires plasmids encoding the flippase recombination enzyme (FLP) to insert into ClosTron mutants. Unfortunately, these plasmid(s) were not available for this study. Further studies with the *cd2791* mutant, any mutants involved in operons, or indeed any KO should include this *ermB* removal step to prevent any misinterpretation of phenotypes.

However, the polar effects of the insertion in CDΔ*cd2791*$_{1569}$ on *cd2790* and *cwp66* transcription may not necessarily affect protein levels. CD2791 is constitutively present on the cell surface, indeed the *cd2791* (*cwp2*) promoter has been used in complementation studies of a CwpV mutant (Emerson *et al.* 2009b). Yet, Cwp66 is

not always found on the cell surface (Wright *et al.* 2005). If, as evidence suggests, *cd2791* and *cwp66* are on the same transcript, finding one and not the other suggests post-transcriptional control mechanism(s). Therefore, regulation of Cwp66 may be performed, not by upregulation of *cd2791-cd2790-cwp66*, but through translational control. Thus, as a transcript is found for *cwp66* in CDΔ*cd2791*$_{1569}$, it is possible that Cwp66 is still translated and (post-transcriptionally) regulated in a wildtype fashion.

The ~50% adhesion of CDΔ*cd2791*$_{1569}$ compared to CDΔ*cwp66*$_{519}$ suggests that even if Cwp66 has been affected, loss of CD2791 also markedly decreases the ability of *C. difficile* to adhere to Caco2 cells. CD2791 is, therefore, putatively an adhesin of *C. difficile*. CD2791 specific inhibitors may putatively further demonstrate the role of CD2791 in adhesion.

5.4.2.4.5 FliD

The adhesion of two mutants appears to be greater than the WT (CD2784 and FliD). The increased adhesion of a FliC mutant has been reported (Baban *et al.* 2010), suggesting that knockout of flagella genes increases adhesion. It is possible that a non-motile, normally motile, bacterium may be in a stressed state and as such, increases stress related (adherence) proteins, e.g. Cwp66. Although, why the other surface protein mutants produced in this chapter, which are also presumably stressed, do not show increased adhesion is unknown. Tasteyre *et al.* (2001) reported that non-flagellated strains did not bind to mouse ceca as well as flagellated strains; one would expect therefore, that flagella mutants should not adhere as well as the flagellated wildtype. In *B. subtilis*, N-acetylmuramoyl-L-alanine amidase (peptidoglycan hydrolase) synthesis is controlled by a regulatory network involving SigD, co-regulated with flagella motility (Lazarevic *et al.* 1992). If a similar network exists in *C. difficile*, where flagella and surface proteins are co-regulated, inhibition of one may change expression of the other. Recently, Jarchum *et al.* (2011) reported that administration of *Salmonella typhimurium* derived flagellin, a TLR5 agonist, protects mice from *C. difficile*. TL5 activation not only decreased colonisation but also prevented epithelia damage. The increased virulence of the

FliC mutant (Baban *et al.* 2010) may result from decreased activation of TLR5 cascades, consistent with TLR5 negative mice being more susceptible to disease (Lawley *et al.* 2009a).

5.4.2.4.6 CD2784 and CD1036

The reasons for increased adhesion of the CD2784 mutant is unknown, especially as this protein displays slower growth than the WT but does not produce any more toxin. It is possible that inactivation of CD2784, results in unnoticed changes to the cell surface resulting in increased intercell adhesion. Increased intercell adhesion would appear to reduce the growth rate as measured by A_{600} and may explain higher adhesion to Caco2 cells (through more bacteria sticking together or sticking to the cells). Further work is needed to confirm the increased adhesion of the CD2784 mutant. The mean adhesion of the CD1036 mutant was not significantly different to the WT, suggesting that CD1036 is not involved in the same pathways as CD2784, although their different genomic locations hint at this.

Peptidoglycan hydrolase mutants typically have normal growth rates but attenuated virulence and may form chains *in vitro* (Sanchez-Puelles *et al.* 1986; Mani *et al.* 1994; Mercier *et al.* 2002). A *Listeria monocytogenes* peptidoglycan hydrolase (IspC) mutant displayed no defects in *in vitro* growth, colony and microscopic morphologies or biochemical characteristics but displayed attenuated in virulence in cell culture models. The attenuated virulence was attributed to the finding that the IspC knockout had altered expression of other surface based adhesins. IspC, CD1036 and CD2784 are similar in that one end is the catalytic domain, while the other is a CWBD. It was proposed that the catalytic domain of IspC alters the peptidoglycan, providing the necessary cell wall architecture for display of other virulence factors (Wang & Lin 2008). The knockout of CD2784 may, therefore, alter the presentation of other surface based adhesins, increasing adhesion. However, the redundancy of peptidoglycan hydrolases (autolysins) makes it difficult to study their interactions.

The absence of a completely non-adherent *C. difficile* mutant highlights how adhesion is likely to be multifaceted process. Moreover, direct relationships to adherence are difficult to infer as the KO of one protein may have dramatic effects on other proteins, and it may be these proteins which affect adhesion. *C. difficile* bacterial adhesion is severely decreased if the surface proteins are removed with 4 M GnHCl or 8 M urea (Takumi *et al.* 1991), suggesting that a relatively limited number of proteins result in bacterial adhesion and subsequent colonisation. If the actions of 4 M GnHCl or 8 M urea could be replicated therapeutically, it may form an effective any anti-colonisation treatment.

5.4.2.5 Adhesion studies of rCD2791$_{27\text{-}322}$

The adhesion assay suggested that the CD2791 mutant had a marked reduction in adherence compared to the WT. To investigate if the CD2791 protein acted as an adhesin directly, fluorescently labelled purified N-terminal domain of CD2791 was incubated with Caco2 and Vero cells.

The binding of rCD2791$_{27\text{-}322}$ to Vero cells at 1 µM suggests that CD2791 is an adhesin. Unfortunately, Vero cells did not adhere well to the glass chamber slides, thus preventing competition studies with unlabelled rCD2791$_{27\text{-}322}$ or further determination of binding affinities. Further work is necessary to confirm the binding of rCD2791$_{27\text{-}322}$ including the instrumentation of confocal laser scanning microscopy (CLSM), which has been widely used to identify bacterial adhesion (including *C. difficile*) to Caco-2 cells, (Miller *et al.* 2000; Cerquetti *et al.* 2002).

Autofluorescence of Caco2 cells has not been reported, therefore it is possible that rCD2791$_{27\text{-}322}$ did not bind to Caco2 cells, and the high exposure gains used, created an artificial impression of autofluorescence. Further work is required to assess binding of rCD2791$_{27\text{-}322}$ to Caco2 cells *in vitro*, including inhibition of *C. difficile* adhesion to cell lines using rCD2791$_{27\text{-}322}$ or anti- rCD2791$_{27\text{-}322}$ antibodies.

Efforts to obtain anti-CD2791 antibodies from anti-R20291 SLP IgG, using rCD2791$_{27\text{-}322}$ coupled to an NHS ester resin, largely failed as only tens of

micrograms of antibodies were purified, too low for use in adhesion studies. Reasons for the low yield include: coupling shielding binding epitopes, CD2791 antibodies being out-competed by S-layer antibodies or that the antibody:antigen interaction is weaker than expected despite, reasonable ELISA titres using sheep anti-R20291 SLP IgG. To further a CD2791 adhesin hypothesis, e.g. prevention of *C. difficile* binding by anti-CD2791, it is necessary for antiserum to be raised in animals.

5.4.3 Summary

In this chapter, a number of genes coding proteins putatively associated with the cell envelope of *C. difficile* have been knocked out using the ClosTron system. Characterisation of the mutants revealed growth rates were largely similar to the WT, apart from the CD2784 mutant. All mutants sporulate, suggesting none of the proteins knocked out are key in forming viable spores. Four mutants demonstrated increased culture supernatant toxin A levels after 24 hr and but demonstrated decreased adherence *in vitro*. The surface protein CD2791 was identified as a putative adhesin, but further work is required to test this hypothesis. Challenge of an animal model with surface protein mutants, may help to determine if *in vitro* results are replicated *in vivo*.

5.4.4 Conclusions

To date, of the 29 proteins in the *C. difficile* genome that contain a PFam04122 binding domain, only 11 have been proven to be expressed and/or surface located. Using a knockout approach, work in this chapter confirms that some of those surface proteins already identified are likely adhesins, e.g. Cwp66, but other, as yet uncharacterised surface proteins may also play a role in adhesion. Therapeutically targeting surface proteins must be approached with caution, as inactivation of several of the surface proteins in this chapter resulted in increased toxin production *in vitro*. The key to progressing knowledge into *C. difficile* adhesion and colonisation remains the identification of specific surface protein(s) with key roles *in vivo* during CDI.

CHAPTER 6 –
Crystallisation of selected *C. difficile* surface proteins

6 CHAPTER 6 – Crystallisation of selected *C. difficile* surface proteins

6.1 Introduction

6.1.1 Introduction to protein crystallography

In order to fully understand protein function, specifically the physical recognition of substrates and any conformational changes induced during action, it is necessary to observe protein structure(s) at atomic resolution. The 3D structural knowledge obtained then provides a basis by which rational drug design and structure based functional studies can develop therapeutics.

To observe protein structure(s) at atomic resolution, two methods are typically used; Nuclear Magnetic Resonance (NMR) (9% known structures) and X-ray crystallography (90% known structures). X-rays are suited to structure determination as X-rays have wavelength in the range of 100 to 0.1 Å (typically 1.5 Å), about the same as the C-C bond length. Crystals of proteins increase diffraction intensities due to their periodicity, i.e. repeated unit cells. Structural determination of proteins using X-ray crystallography ultimately obtains structures of proteins by back calculation of the physical locations of the atoms from the resulting protein crystal X-ray diffraction pattern (electron density). The typical process is outlined in Figure 6.1.1.

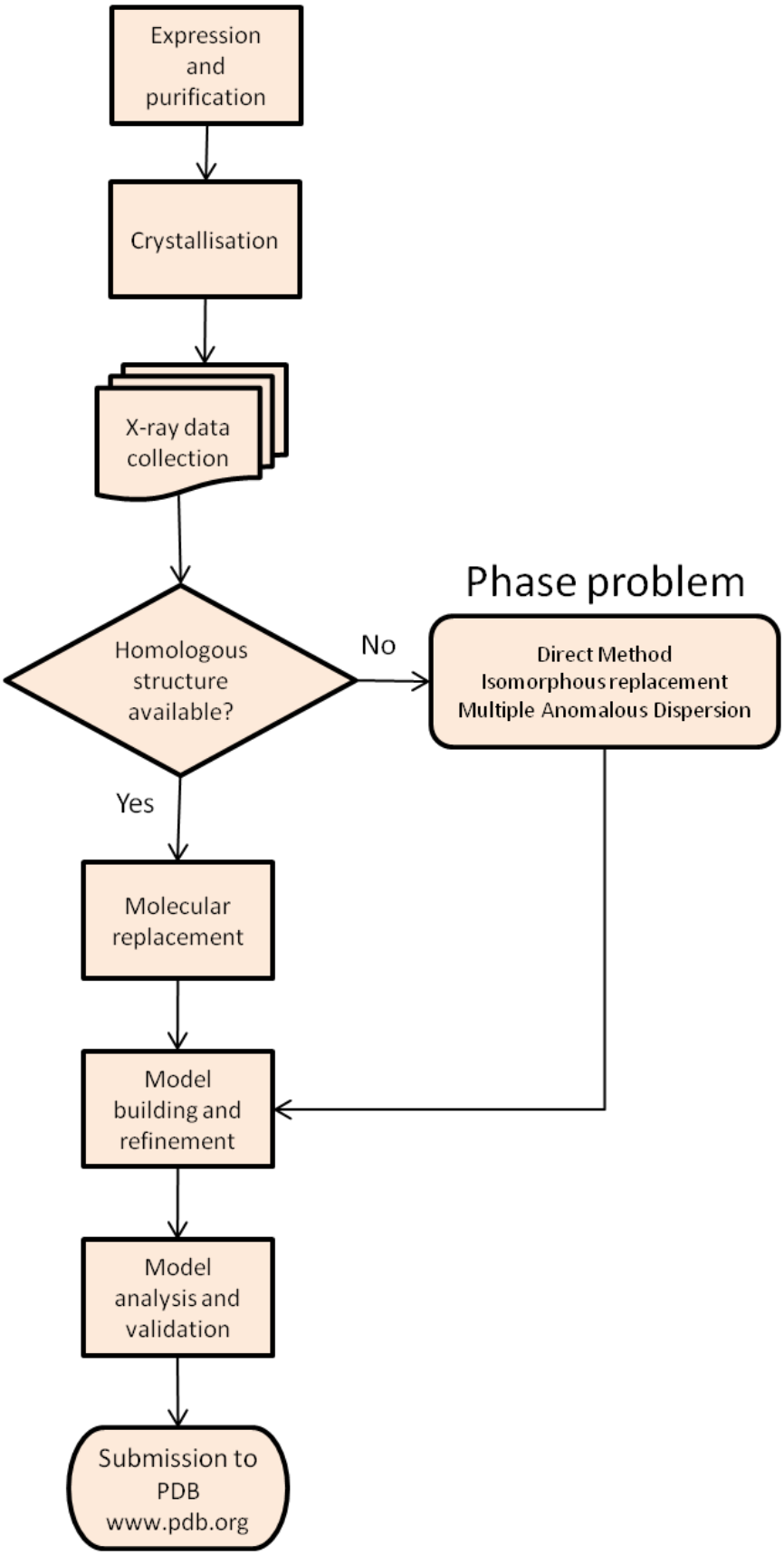


Figure 6.1.1 Outline of steps involved in protein crystallography

6.1.1.1 Expression and purification

The goal of this step is to obtain pure (>90%) protein, as any contamination of the protein sample could affect the way that the protein crystallises, if the sample crystallises at all. The ideal source of protein would be the native source, for example, a *C. difficile* surface protein extract. Commonly, to increase yield, the chosen protein is over-expressed recombinantly. The benefits of recombinant expression are many fold, e.g. inclusion of tags, modification/expression of domains, with an aim to produce soluble protein, although the expression system often asserts its own restrictions on that aim. In most cases, however, these restrictions can be overcome.

Once protein is obtained it is often necessary to remove any contaminating proteins. These may be natively derived or as a result of the recombinant expression system. Several different purification systems exist and are detailed in several texts (particularly those by GE Healthcare). Protein purification relies on interactions of the protein with various chromatographic media. This interaction can be based on size (gel filtration), charge (ion exchange), specific protein interaction (affinity), or hydrophobicity (hydrophobic interaction or reverse phase). One or more of these may be employed in order to obtain pure protein, or, at least, as pure as possible to attempt crystallisation.

6.1.1.2 Crystallisation

While the process of crystallisation is well understood, the determination of conditions necessary for protein crystallisation is largely empirical (verging on luck), often requiring screening of hundreds of crystallisation conditions. Crystallisation occurs when the protein concentration in solution is greater than its limit of solubility, thus the protein enters a supersaturated state (Figure 6.1.2).

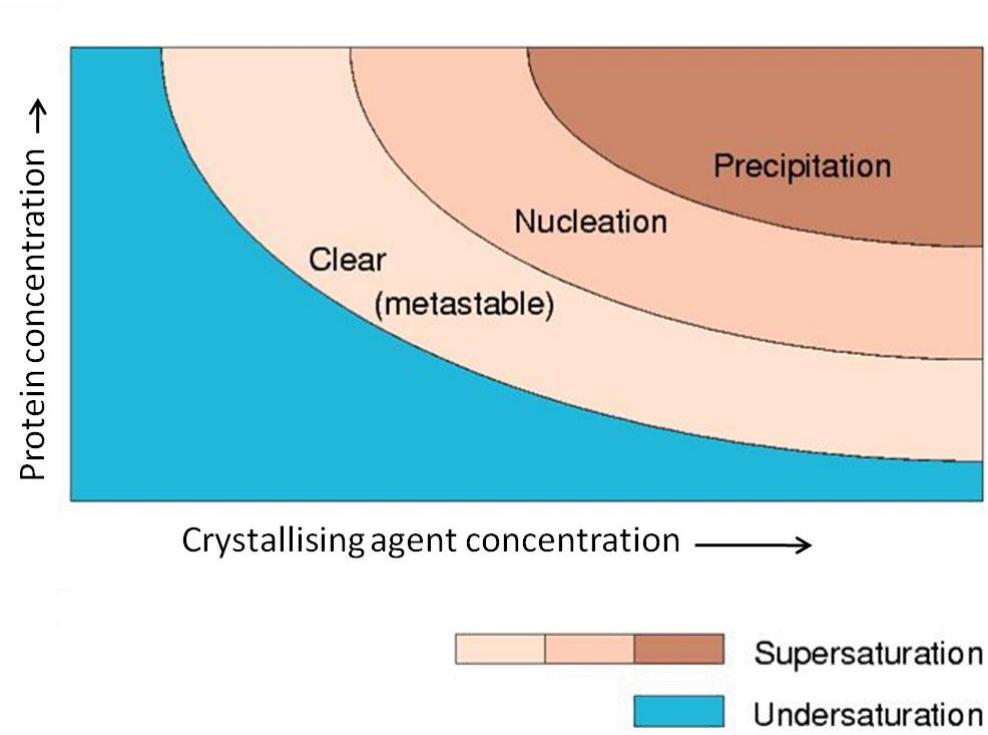


Figure 6.1.2 Phase diagram for vapour diffusion. Adapted from McCoy (2005)

Once the protein is in a suitable state, nucleation and subsequent crystal growth occurs in three dimensions. As proteins are irregularly shaped, this results in the formation of large, solvent filled, channels in the crystals. Protein crystals are therefore crystalline solids, an intermediate between a fluid and a solid held together by non-covalent interactions. These solvent channels can be used to the crystallographer's advantage, where crystals can be soaked with a solution to study ligand interactions or with heavy metals to solve the phase problem (discussed later in 6.1.1.5).

Physically crystallising a protein can be accomplished by several methods including dialysis, batch and vapour diffusion. The most common method of crystallisation is vapour diffusion (sitting drop or hanging drop), in which solvent in a protein drop (plus a small amount of reservoir solution) and reservoir solution equilibrate by vaporisation, resulting in an increase of the precipitant concentration in the drop to a level optimal for crystallisation of the protein.

6.1.1.3 X-ray data collection

The obtained crystals are placed in a beam of X-rays; the resulting diffraction of the X-ray beam produces (Laue) spots (reflections) on a detector, Figure 6.1.3.

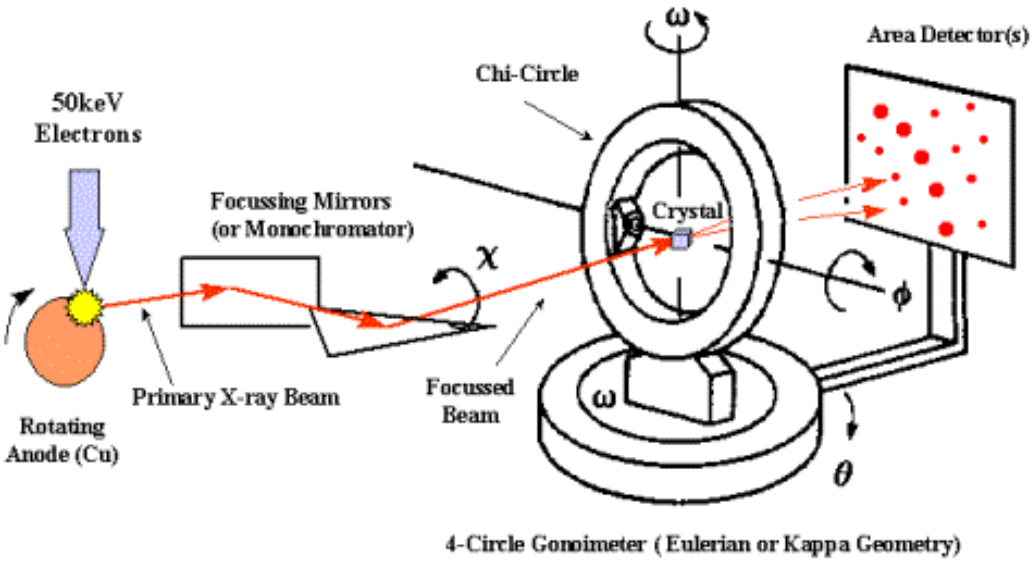


Figure 6.1.3 X-ray diffraction set-up. Taken from http://pruffle.mit.edu/atomiccontrol/education/xray/xray_diff.php.

The crystal is then rotated to collect a full data set. The amount of rotation need not be 360°, but is largely determined by the mathematical concept of Ewalds sphere (Ewald 1969) and the crystal symmetry (space group) (higher symmetries need less rotation due to the likelihood of seeing the same plane). Crystals are made up of several families of planes (called lattice planes or Bragg's planes), thus using simple wave mechanics: if reflections from a crystal are in phase (constructive interference) a spot is observed, if they are not, they cancel out (destructive interference). The angle and distance between lattice planes can be derived from Braggs Law:

$$n\lambda = 2d\sin\theta$$

Where *n* is an integer (in constructive interference), λ is the wavelength of a beam of incident x-rays on a crystal with lattice planes separated by distance *d,* and θ is the Bragg or scattering angle (illustrated in Figure 6.1.4).

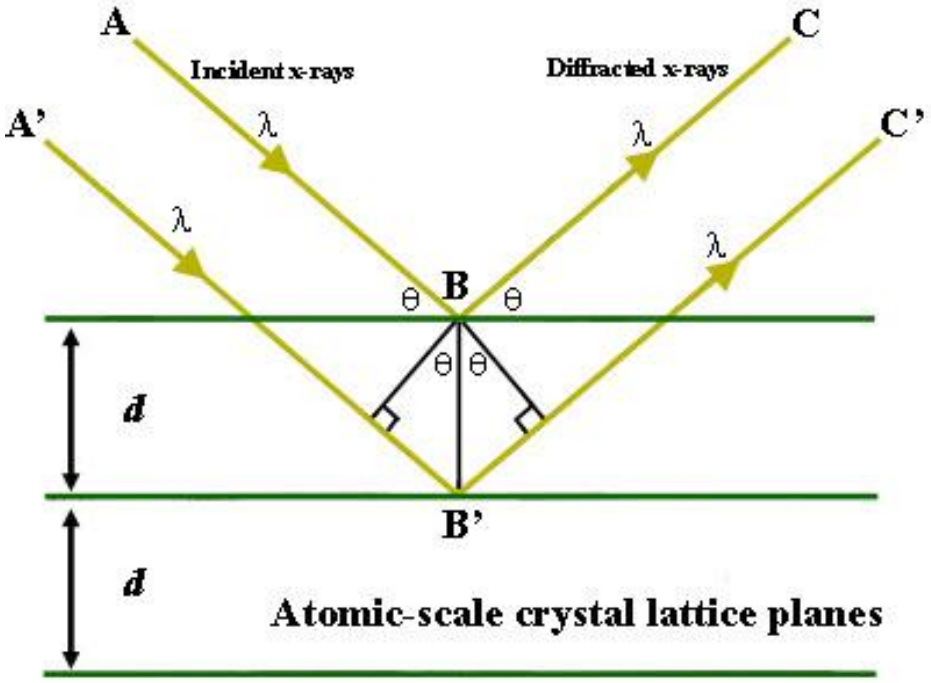


Figure 6.1.4 Diagrammatic representation of Braggs Law. Taken from http://serc.carleton.edu/research_education/geochemsheets/BraggsLaw.html

Protein crystals deconstructed to their simplest form are made up of asymmetric units (Figure 6.1.5). By definition, this is the smallest portion of a crystal structure to which crystallographic symmetry can be applied to generate one unit cell. There may be several protein molecules per asymmetric unit, or just one. A unit cell is the element repeating that makes up the crystal in three dimensions. The unit cell can be described by the lengths of three edges (a, b, c) and the angles between them (α, β, γ) (Figure 6.1.6).

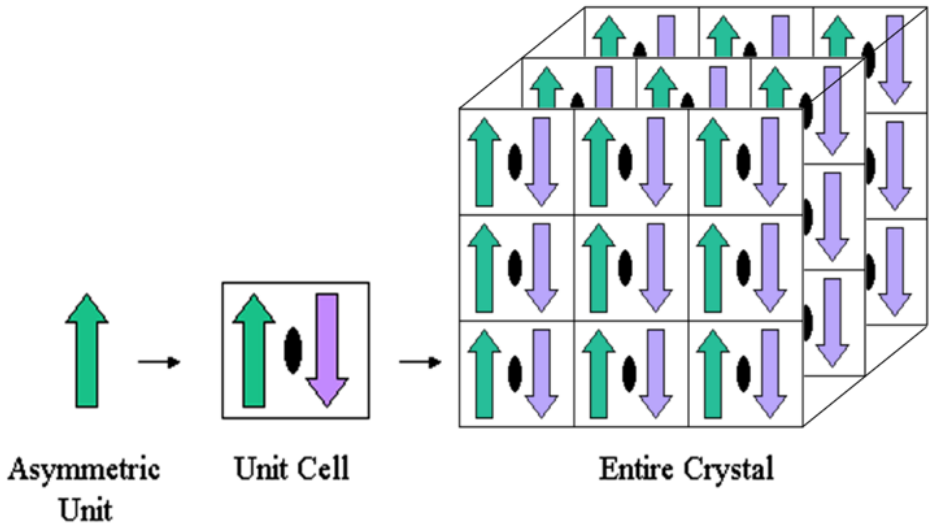


Figure 6.1.5 Fundamental parts of crystal structures. The asymmetric unit (green upward arrow) is rotated 180 degrees about a two-fold crystallographic symmetry axis (black oval) to produce a second copy (purple downward arrow). The asymmetric unit may contain several protein molecules or just one. Together the two arrows comprise the unit cell. The unit cell is then translationally repeated in three directions to make a 3-dimensional crystal. Taken from http://www1.rcsb.org/pdb/static.do?p=education_discussion/Looking-at-Structures/bioassembly_tutorial.html

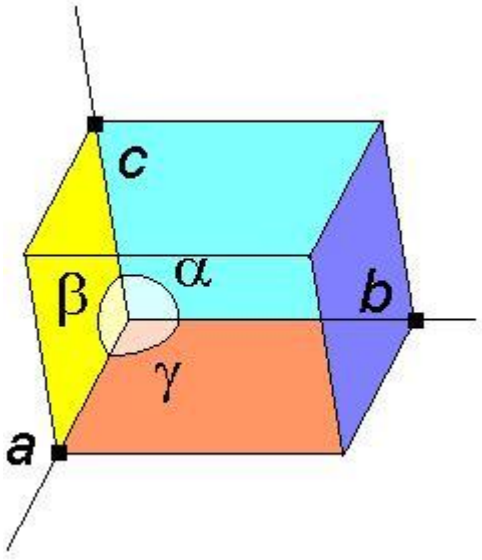


Figure 6.1.6 Measurable parameters of crystals

As all six of these dimensions can be varied, it gives rise to seven lattice systems with an increasing order of symmetry based on the number of parameters set (Cubic) to no parameters set (Triclinic) (Table 6.1.1). In addition to these seven primitive lattices, Bravais discovered another seven non-primitive lattices with additional lattice points at the centre of the cell (I), on the centre of each face (F), or on only one face (A, B or C) (a total of 14 Bravais lattices are thus possible).

Table 6.1.1 Types of Bravais lattices

Crystal system	Number of Lattices	Lattice symbols	Parameters
Triclinic	1	P	a≠b≠c α≠β≠γ
Monoclinic	2	P, C	a≠b≠c α=β=90° ≠ γ
Orthorhombic	4	P, C, I, F	a≠b≠c α=β=γ=90°
Tetragonal	2	P, I	a=b≠c α=β=γ=90°
(Rhombohedral) Trigonal	1	P	a=b=c α=β=γ≠90°
Hexagonal	1	P	a=b≠c α=β=90° γ= 120°
Cubic	3	P, I, F	a=b=c α=β=γ=90°

6.1.1.4 Data analysis and model building

The two dimensional diffraction images must then be converted to a 3D representation of the electrons which diffracted the X-rays. To do this, firstly, the images are indexed, i.e. identifying the dimensions of the unit cell and which image peak corresponds to which position in reciprocal space. The term 'reciprocal space' is used as there is a reciprocal relationship between the distances between lattice planes (*d*) and the scattering angle (θ). Smaller distances cause wider diffraction

patterns in ‘real’ space, whereas lattices with a larger d result in a closer diffraction pattern.

The intensity of the spots from the images is then integrated and the complete data merged and scaled into one data file. This process resolves which peaks appear in two or more images (merging) and scales the relative images so that they have a consistent intensity scale.

Using an inverse Fourier transform (as crystals are periodic), the electron density can be derived from the structure factor equation. To fulfil this equation and calculate the electron density, there is one piece of information missing, the phase (lost through the physical act of measuring the diffracted waves). To put it another way, the obtained x-ray data contains the structure factor magnitudes in the form of intensities/amplitudes but not their positions.

6.1.1.5 The phase problem

There are several methods in which to obtain the phase or at least estimate the phase:

- The Direct Method

Typically applied to molecules with <1000 atoms per asymmetric unit or if the resolution is <1.2 Å. Individual atoms can be fitted and the problem is join-the-dots.

- Molecular replacement (MR)

When the protein being studied shares >20% identity with a known structure in the PDB, over a reasonable portion of the protein, the known molecular model is used to solve the unknown crystal structure. The known model is orientated and positioned such that the predicted diffraction (as if the known model had crystallised) best matches the observed diffraction (Evans and Coy, 2008). MR assumes that proteins with similar amino acid sequences share tertiary structure; this is not always the case. Homology models used as templates may work in MR but on the same basis as primary sequence identity, i.e. are good matches without

large sequence deletions/additions. Using a known model can however bias the model of the unknown protein.

- Multiple Isomorphous Replacement (MIR)

Crystals are soaked in heavy metal solutions, such that the heavy metals diffuse into spaces originally occupied by the solvent. The heavy metal should be introduced such that the crystal lattice is unaltered, i.e. isomorphous. By comparing the reflections generated by several different isomorphous crystals, the positions of the heavy atoms can be worked out and this allows the phase diffraction in the un-substituted crystal to be deduced.

- Multiple Wavelength and Single Wavelength Anomalous dispersion (MAD and SAD)

Similar to MIR, heavy atoms are incorporated into proteins, usually by substituting sulphur for selenium, in methionine (selenomethionine) (or sulphur for cysteine in selenocysteine (Strub *et al.* 2003)). This method is applicable for native metallo-proteins or for recombinant proteins, and requires approximately one selenomethionine residue for every ~75–100 amino acids (Hendrickson *et al.* 1997; Strub *et al.* 2003). By changing the X-ray wavelength from where the metal does not absorb (normal diffraction, obeying Friedels law $F_{hkl} = F_{hk}$), to where the metal does absorb (anomalous diffraction, disobeying Friedels law $F_{hkl} \neq F_{hk}$) and comparing the intensities of Friedel's pairs, the position of heavy atoms in the unit cell are determined.
Combinations of the two heavy metal methods can be applied, utilising the benefits of both e.g. Single wavelength Isomorphous Replacement with Anomalous Scattering (SIRAS) and Multiple wavelength Isomorphous Replacement with Anomalous Scattering (MIRAS).

6.1.1.6 Refinement, Validation and Submission

When a model has been built, it is refined, i.e. adjusted to improve the agreement with the measured diffraction data. Refinement modifies bond length, bond angle,

torsion (dihedral) angle, and planarity of the peptide bond together (Figure 6.1.7) with the thermal vibration 'temperature' (B-factor), to give a better fit for each atom.

From this, a number called the reliability factor (R-Factor, or R_{work}) is calculated which is simply a measure of the agreement between the crystallographic model and the experimental X-ray diffraction data (the lower the better with typically 0.2 for well refined protein model). However, it is possible to over fit or 'misfit' the diffraction data. By using a randomly selected (typically ~10%) set of reflections that are omitted in the modelling and refinement process, it is possible to measure the agreement between observed and computed structure factor amplitudes (R_{free}) (Brunger 1992).

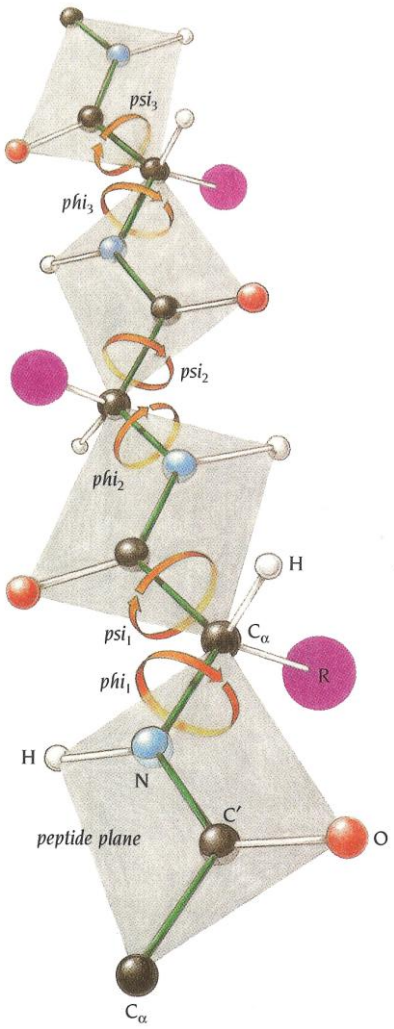


Figure 6.1.7 Diagrammatic representation of polypeptide chain and the degrees of freedom. Rotation can occur along the C_α- C' bond (psi ψ) or N-C_α bond (phi ϕ). Taken from Branden & Tooze (1999).

After refinement, one can assess, i.e. validate, how successful the model is by the use of a number of statistical or graphical figures. A common assessment is the Ramachandran plot which assesses the phi and psi torsion angles to identify amino acids of the polypeptide backbone in sterically disallowed conformations. Once the structure has been validated it is submitted to the Research Collaboratory for Structural Bioinformatics (RCSB) protein data bank (PDB) (www.pdb.org).

6.1.2 Proteins selected for crystallisation

The bacterial cell surface serves two central purposes: protection from and interaction with, the environment. Infections by microbes follow a similar pattern: adhesion to host cells followed by complex interactions that often involve secretion of effector proteins, e.g. toxins and induction of signalling processes by the host cells. The proteins of the cell surface are of considerable interest both academically and industrially as they represent a novel interaction site between host and pathogen. The interruption of such interactions may result in the cessation of a potentially harmful infection.

The sequence of amino acids which make up (surface) proteins can provide a great deal of information about the likely roles of the protein, primarily in relation to (and as a result of) other already characterised proteins. It is the tertiary or 3D structure of the protein that reveals the true interactions between proteins and their ligand/receptor interactions. Understanding this complex interaction requires data at the molecular or even atomic level obtained by X-ray crystallography and/or Nuclear Magnetic Resonance spectroscopy (NMR).

C. difficile proteins, for which structural information has been obtained, has understandably been directed towards the main virulence factors: the toxins (Ho *et al.* 2005; Pruitt *et al.* 2009; Sundriyal *et al.* 2009; Albesa-Jove *et al.* 2010; Pruitt *et al.* 2010; Puri *et al.* 2010).

To date, there is only one report of high-resolution structural information for a *C. difficile* surface protein, a fragment of the LMW SLP (3CVZ) (Fagan *et al.* 2009) (see General Introduction 1.2.2) and a structure submitted, but no published literature, for a putative membrane protein (CD3042)(3KMI) (Chang *et al.* 2009). Moreover, given the large number of the 28 slpA paralogs which have yet to be assigned even a hypothetical function, there are considerable gaps in the understanding how the surface proteins of *C. difficile* interact with both themselves and the environment.

Structural data is likely to be key in furthering our understanding of S-layer biogenesis, the interactions of the CWBD with the underlying cell surface and how surface proteins interact with themselves or host cell proteins.

To begin to address this issue a number of surface proteins containing the CWBD (PFam04122) were selected for obtaining x-ray crystallographic information. The determination of function from structure was the primary goal in this chapter, particularly trying to obtain structural information on full length surface proteins, including their CWBDs. It was discovered that the lack of homology to known proteins prevents all but speculative function prediction and it is likely that only x-ray crystallographic structural information will reveal the folds (novel or not) to help identify functions for these important proteins.

6.1.2.1 CD2791 (Cwp2/ORF2, ORF8)

623 aa, (66,445Da predicted)

Signal peptide	1-26
N-terminus	Unknown function
2x CWBD	326-417, 426-512

Figure 6.1.8 Domain structure of CD2791. Sig- Signal Peptide, CWBD- Cell wall binding domain. Putative domain location(s) obtained from *C. difficile* 630 genome annotations, available at ncbi.nlm.nih.gov.

The *cd2791* gene was first identified in 2001 together with the identification of the cluster of genes surrounding slpA (Calabi *et al.* 2001; Karjalainen *et al.* 2001).

The *cd2791* gene is highly conserved in all strains and appears to be transcribed during normal growth (Calabi *et al.* 2001) and is suggested to be at the start of a polycistronic message also encoding CD2790 and Cwp66, putatively part of an adhesin complex (Savariau-Lacomme *et al.* 2003). However, this complex has not been found on the cell surface to date. Microarray analysis found one of duplicate

reporters for *cwp2(cd2791), cwp84, cwp6(cd2784)* and *cwp7(cd2782)* was upregulated during exposure to sub-MIC amoxicillin (Emerson *et al.* 2008).

The protein encoded by *cd2791* was initially described as an approximately 70 kDa protein, identified on the cell surface in guanidine SLP extracts of all isolates by McCoubrey & Poxton (2001). The authors also reported that this protein cross-reacted with all antisera tested, consistent with data in Chapter 4. A 66 kDa protein on the cell surface was also reported by Calabi & Fairweather (2002), who demonstrated that it had lost its signal peptide and did not appear to undergo post-translational modification as the predicted MW matched the observed. Moreover, N-terminal sequencing of the 66 kDa protein matched the N-terminus of Cwp2 (ORF2) (CD2791) in the unfinished *C. difficile* 630 genome. BLASTP of CD2791 revealed homologies to both the S-layer of *Bacillus sphaericus* and the transducer of rhodopsin (Htr) II from the archaebacterium *Natronobacterium pharaonis* (Calabi & Fairweather 2002). Proteomic analysis of low pH or lysozyme cell wall extracts confirmed that the 66 kDa protein found in surface extracts was CD2791 (Wright *et al.* 2005) and perhaps unsurprisingly, given its apparent constitutive expression, CD2791 is also found in the spore coat (Lawley *et al.* 2009b). However, despite being surface located, an N-terminal 38-41 kDa protein fragment of CD2791 appears to be present in the culture supernatant, particularly during conditions promoting high toxin production (Mukherjee *et al.* 2002).

Secondary structure analysis suggested that the N-terminal, predicted 'functional' domain of CD2791, consisted of a largely alpha-helical conformation (71%) with a putative trans-membrane segment, which shared no significant homology with known proteins (Savariau-Lacomme *et al.* 2003).

As discussed in Chapter 4, CD2791 appears to be highly immunogenic as virtually all patients raise antibodies to CD2791 (Wright *et al.* 2008). This, coupled with a lack of CD2791 in strain 167 (a clinical isolate) (Calabi & Fairweather 2002), highlights how antibodies to CD2791 are likely to be non-protective and that CD2791 does not appear to be necessary *per se* for CDI. However, the altered S-layer and lack of

CD2791 in strain 167, may be negated by changes in other virulence associated factors e.g. increased adhesion or increased toxin production, particularly as the Cwp84 mutant (with an immature S-layer) is still able to cause CDI in the hamster model (Kirby *et al.* 2009).

Therefore, CD2791 represents a good protein for which to undertake structural studies, as CD2791's constitutive expression, immunogenicity and presence on the cell surface suggest it is a key protein to *C. difficile* either physiologically or pathogenically. Data in Chapter 5 also suggest CD2791 may play a role in adhesion, further supporting the precedent to obtain structural information about this protein.

6.1.2.2 Cwp66 (cwp3 (Calabi *et al.* 2001)

610 aa (66,779Da predicted)

Signal peptide	1-29
3x CWBD	33-121, 130-221, 228- 322
C-terminus	Unknown function

Figure 6.1.9 Domain structure of Cwp66. Sig- Signal Peptide, CWBD- Cell wall binding domain. Putative domain location(s) obtained from *C. difficile* 630 genome annotations, available at ncbi.nlm.nih.gov.

A putative heat shock associated cell wall protein of 66 kDa was first characterised in 2001 (Waligora *et al.* 2001), after initial data suggested that heat shock increased adherence in *C. difficile* (Eveillard *et al.* 1993). Analysis of *cwp66* expression by RT-PCR confirms that *cwp66* expression occurs during stress conditions, e.g. heat shock, as *cwp66* was upregulated during high osmolarity and showed a slight increase in the presence of sub-MIC concentrations of ampicillin, but not iron limiting conditions (Deneve *et al.* 2008).

Waligora *et al.* (2001) also suggested that *cwp66* was unlikely to have its own promoter since no prominent promoter structure could be identified in the short intergenic region between *cd2790* (*orfA*) and *cwp66* (*orfB*). Savariau-Lacomme *et al.*(2003) suggests *cwp66* is encoded on a polycistronic message with *cd2790* and *cd2791*, despite *cd2791* (ORF8) (and *cd2790* (ORF9)) being preceded by promoter consensus sequences at -10 and -35. Savariau-Lacomme *et al.*(2003) also suggests that expression of the polycistronic message was found at only the beginning of exponential phase. However, given the constitutive presence of CD2791 on the cell surface, this may suggest how transcription may be linked but translation is not. The ability of *cwp66* to be inhibited by its own antisense mRNA was assessed, but no significant effect on Cwp66 protein levels or adhesion *in vitro* was observed (Roberts *et al.* 2003).

Cwp66 contains three N-terminal CWBDs and a C-terminal domain which is predicted to have a secondary structure of an extended beta strand formation. Using immunogold labelling anti-*cwp66* N- or C- terminal antibodies did not detect *cwp66* on the surface of *C. difficile* grown in normal conditions, however, after cells were heat shocked, anti-*cwp66* C-terminal antibodies reacted strongly suggesting that the C-terminal domain was surface most, at least after heat shock. Adhesion to Vero cells was prevented by recombinant *cwp66* N- or C-term at 50 µg/ml and anti-*cwp66* antibodies prevented adhesion to ~50%, however, only when heat shocked bacteria were used (Waligora *et al.* 2001).

There appears to be some debate as to the existence of Cwp66 on the cell surface during normal conditions, as Cwp66 (aka paralog 3) was not found on the cell surface during proteomic analysis of cell wall extracts (using low pH or lysozyme extraction of un-heat shocked cells (Wright *et al.* 2005)). Whereas Cwp66 was used in two studies as a cell wall (peptidoglycan and SLP fraction) protein marker, specifically detected by anti-Cwp66 C-terminal antiserum (Hennequin *et al.* 2001; Hennequin *et al.* 2003). Cwp66 has also been found to co-purify with ABP labelled Cwp84 (together with CD2767 and CD2797), suggesting that it may form part of a complex on the cell surface (Dang *et al.* 2010).

As discussed in Chapter 4, the large majority of the surface proteins are highly conserved whereas the C-terminal domain of Cwp66 is highly variable and three different 'typing' groups have been suggested (Waligora *et al.* 2001; Calabi & Fairweather 2002).It is suggested that *cwp66*, like *slpA*, may be under strong environmental selective pressure (Lemee *et al.* 2005). Karjalainen *et al.* (2001) also suggests that the variable domain of Cwp66 is up to 40% homologous with the 'variable' domains of *Orf10(cd2786), Orf5(cd2795), Orf6(cd2794)* and *Orf2(cd2798)*, and that the third CWB repeat is less conserved than the first two.

During CDI, the C-terminal domain of Cwp66 appears to generate a greater antibody response in a greater number of patients (14/17), than the N-terminal domain (02/17) (Pechine *et al.* 2005b), supporting Waligora *et al.* (2001) suggesting that the C-terminal is surface exposed. Although, analysis of CDI case patient serum by Pechine *et al.* (2005a) found the anti-Cwp66 C-terminal response was not statistically different from controls, whereas the lower anti-Cwp66 N- terminal response appeared to be statistically significantly higher in cases than in controls. However, the control serum used by Pechine *et al.* (2005a) was from maternity wards, which may have skewed due to the high culture rates of *C. difficile* from infants and the likely presence (and possible pre-exposure) of *C. difficile* in the environment. Identification of immunoreactive proteins in SLP extracts using CDI patient serum did not detect Cwp66 (Wright *et al.* 2008) although this may have been due to a lack of this protein in the SLP extract used (Wright *et al.* 2005).

Therefore, Cwp66 marks one of the first identified adhesins of *C. difficile* particularly after heat shock. How transferable heat shock results are to *in vivo* conditions, during CDI, is unknown but a change in surface presentation, e.g. adhesins, as a result of stress/environmental conditions, is feasible. Similar to CD2791, Cwp66 also lacks homology or similarity with known proteins, structural determination may therefore aid understanding of this *C. difficile* adhesin. The adhesive nature of Cwp66 may lend itself to ligand docking experiments and/or determination of host receptors.

6.1.2.3 CD2767

703 aa (77,829Da predicted)

Sig peptide	1-26
N-terminus	27-371 DUF187 (Uncharacterised BCR, COG1649; pfam026387)
3x CWBD	402-493, 505-596, 609-684

Figure 6.1.10 Domain structure of CD2767. Sig- Signal Peptide, DUF187- PFam02638, CWBD- Cell wall binding domain. Putative domain location(s) obtained from *C. difficile* 630 genome annotations, available at ncbi.nlm.nih.gov.

The *cd2767* gene is downstream of a large cluster of capsular polysaccharide synthesis proteins and polysaccharide glycosyl transferases (*cd2778* - *cd2764*). To date, CD2767 has been identified as cell wall protein recognised by human antisera to surface proteins (Wright *et al.* 2008) and during a pull down assay of ABP labelled Cwp84 (Dang *et al.* 2010). Publications examining the cell surface and cell surface extracts proteomically (Wright *et al.* 2005) have not found CD2767, implying that either it is compartmentalised such that existing cell wall extraction techniques cannot removed it, its surface based location only occurs in specific conditions, e.g. *in vivo* or upon the inhibition of Cwp84, or the protein is extremely labile and thus not easily recovered.

CD2767 has (like SlpA and Cwp84) three C-terminal CWBD and a 'functional' DUF187 N-terminal domain. Although DUF187 specifically refers to a protein domain of unknown function, protein signature searches suggest the DUF187 family is a member of TIM barrel glycosyl hydrolase clan, which contains members such as alpha-amylase, cellulase and a variety of glycosyl hydrolase families which in total contains 56,730 proteins.

DUF187 is found in diverse species including the YngK protein of *Bacillus subtilis* and *Bacillus amyloliquefaciens* FZB42 and the FenI protein of *Streptosporangium roseum* DSM 43021. In *B. subtilis*, YngK is suggested to be involved in the synthesis of an antifungal lipopeptide antibiotic, pliplastin (Tsuge *et al.* 1999). Inactivation of YngK, results in upregulated sigma X (Turner & Helmann 2000), an extracytoplasmic sigma factor that functions in regulating cell envelope modification as a defence against cationic antimicrobial peptides (Cao & Helmann 2004). A gene in *C. difficile* strains NAP8 and NAP7 has been labelled as YngK, in both strains the gene is 99% similar to CD2767 of *C. difficile* 630. This gene of *C. difficile* strains NAP8 and NAP7 genes are annotated in Uniprot as possessing putative carbohydrate metabolic activity, but also having cation binding activity, i.e. interacting selectively and non-covalently with cations (charged atoms or groups of atoms with a net positive charge) suggesting that cations may be required by CD2767 in some way.

The TIM barrel is a characteristic protein fold named after triose phosphate isomerase (TIM) which contains eight α-helices and eight parallel β-strands arranged such that a central 'barrel' is formed (Wierenga 2001) (Figure 6.1.11).

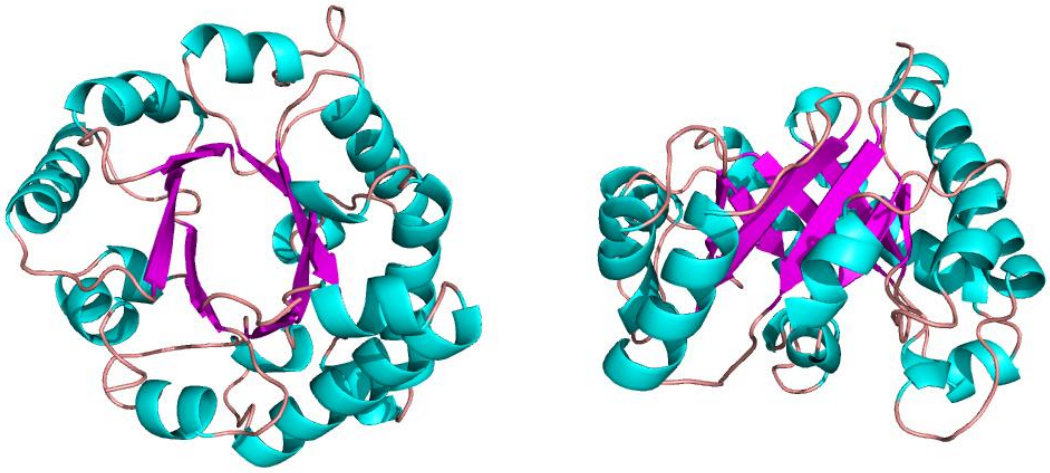


Figure 6.1.11 The structure of the TIM barrel. Top (left) and side (right) views of triosephosphateisomerase (TIM) (8TIM). Prepared using PyMOL (www.pymol.org).

The loops at the C-terminal end of the 'barrel' sheets tend to contain the active site. The α/β barrel (TIM barrel) is by far the most common tertiary fold observed in high resolution protein crystal structures and it is estimated that 10% of all known enzymes have this domain (Farber & Petsko 1990).

Glycosyl hydrolases (glycosidases or glycoside hydrolases) hydrolyse the glycosidic bond between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety. The fold is more conserved than the sequence; as such nomenclature is now based on sequence homology and structural comparisons, complied in the Carbohydrate-Active Enzymes database (CAZy) (Cantarel *et al.* 2009).

Hydrolysis of the glycosidic bond requires two critical residues: a proton donor and a nucleophile/base (commonly glutamic acid (E) or aspartic acid (D)). Hydrolysis occurs via two major mechanisms, depending on spatial location of the nucleophilic catalytic base, as the position of the proton donor is identical in both.

The catalytic base is in close vicinity of the sugar (5.5 Å) in retaining enzymes, whereas the accommodation of a water molecule results in a greater (10 Å) distance in inverting enzymes. Retaining glycosidases operate through a two-step mechanism, with each step resulting in inversion, for a net retention of stereochemistry (Figure 6.1.12A). Inverting glycosidases result in inversion of anomeric configuration, this is generally achieved via a one step, single-displacement mechanism involving oxocarbenium ion-like transition states (Figure 6.1.12B). Active site surface topology determines substrate specificity, as pocket-like active sites are found in exoenzymes, which cleave monosaccharide units from chain ends, whereas endoenzymes have active sites situated in clefts on enzyme surfaces, allowing cleavage to occur in the middle of the chain (Davies & Henrissat 1995). Known inhibitors are sugar shaped molecules e.g. swainsonine, castanospermine and the antiviral Tamiflu (oseltamivir).

Therefore, CD2767 represents one of the few *C. difficile* proteins whereby a function can be postulated. Structural determination of CD2767 may therefore confirm this function i.e. glycosidase and allow further investigations into its role in *C. difficile*.

6.1.2.4 Aim

In summary, the genes chosen for recombinant expression and crystallisation consist of:

- An immunogenic conserved cell surface protein (CD2791)
- A variable cell surface based heat shock inducible adhesin (Cwp66)
- A putative cell surface based TIM barrel glycosidase (CD2767)

Both Cwp66 and CD2791 contain a domain (C-term in Cwp66 or N-term in CD2791) which appears to have little similarity to known proteins using protein signature searches, thus an opportunity exists to gain a structural insight into these proteins which may reveal function from structure. Similarly for CD2767, structural determination will hopefully provide information as to specific positioning of active site residue(s), which often determines glycosidase substrate/ligand binding and specificity.

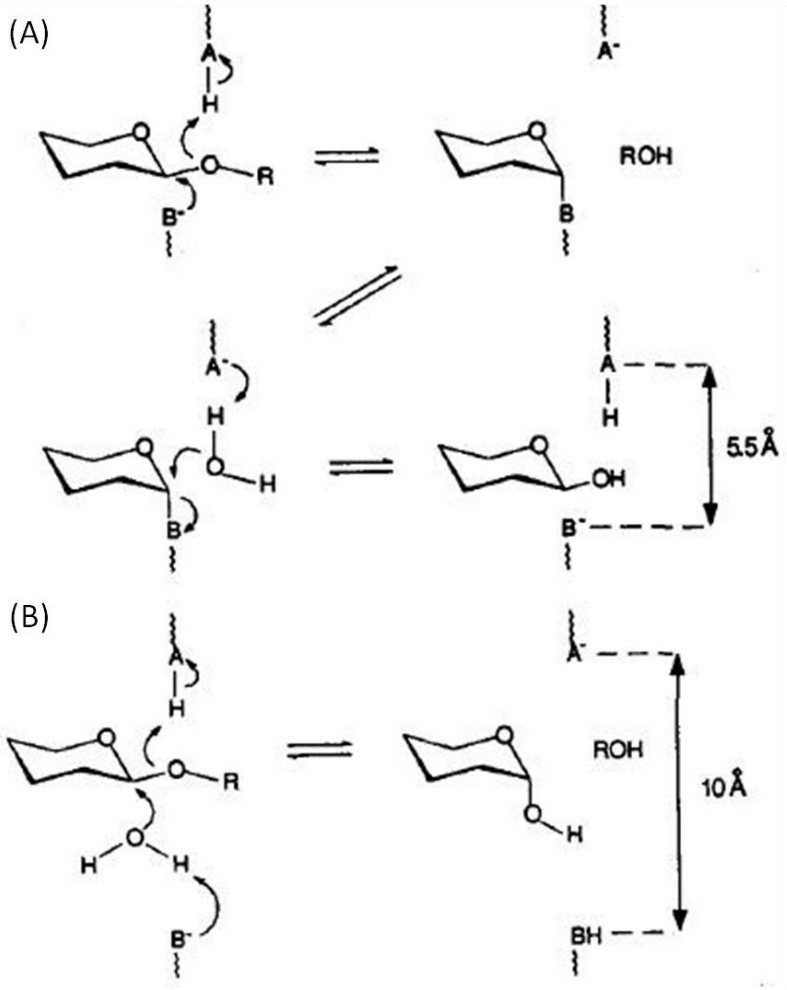


Figure 6.1.12 The two mechanisms of glycoside hydrolases (A) Retaining – In the first step, the glycosidic oxygen is protonated by the acid catalyst (A-H) and nucleophilic assistance to aglycon departure is provided by the base B-. The resulting glycosyl enzyme intermediate is hydrolysed by a water molecule and this second nucleophilic substitution at the anomeric carbon generates a product with the same stereochemistry as the substrate. (B) Inverting – involves a single-displacement mechanism involving oxocarbenium ion-like transition states where protonation of the glycosidic oxygen and aglycon departure are accompanied by a concomitant attack of a water molecule that is activated by the base residue (B-) yielding a product with opposite stereochemistry to the substrate. Figure taken from Davies & Henrissat (1995).

6.2 Chapter Specific methods

6.2.1 Genetic methods

6.2.1.1 Synthetic synthesis of full length *C. difficile* 027 (R20291) strain *cwp66*, *cd2791* and *cd2767* genes

C. difficile 630 gene sequences were BLAST searched against the Sanger Institute's *C. difficile* strain R20291 (Stoke Mandeville, UK) assembly (*cwp66*) or Pathema's QCD32g58 (Quebec, Canada) complete genome (*cd2791* and *cd2767*) (March, 2008). The BLAST result match was translated *in silico*, and the resulting the amino acid sequence (minus predicted signal sequences), flanked by BamHI and XhoI restriction sites followed by two N-terminal stop codons, was submitted to GENEART for *E. coli* optimised gene synthesis and direct insertion into pET28(a) using the same two restriction sites.

6.2.1.2 Insertion of optimised surface protein genes into a range of expression vectors

The synthetically synthesised genes were cloned into a number of expression vectors, assessed to investigate different fusion tags (Table 6.2.1).

Table 6.2.1 Expression vectors used in this chapter

Vector Details / **Gene**	**pET28a** (Novagen) N-term His.T7 kan^R	**pET43.1a** (Novagen) N-term NusA.His amp^R	**pGEX-6P-1** (GE Lifescience) N-term GST amp^R
cd2791	✓	N/D	N/D
cwp66	✓	N/D	✓
cd2767	✓	✓	✓

N/D not determined

6.2.2 Expression of full length selected surface proteins

6.2.2.1 General expression protocol

Constructs were expressed in *E. coli* BL21 (DE3) Star™ as per the large scale expression protocol, General Methods (Section 2.3.3)

6.2.2.2 IMAC Purification

IMAC was performed as detailed in General Methods (Section 2.4.1.3)

6.2.2.3 IEX and HIC purification

IEX was performed as detailed in General Methods (2.4.2). IEX (Q sepharose) was performed as described at pH 7.5 or pH 8.0, whereas IEX (S sepharose) was performed at only pH 5.5. HIC was performed using 5 ml HiTrap phenyl HP (Section 2.4.3) using start/binding buffer- 50 mM sodium phosphate, 1 M NaCl pH 7.0 final buffer 50 mM sodium phosphate pH 7.0, elution of pure (>90%) His.T7.CD2791 was performed using H_2O. Fractions were stored at 4 °C, then pooled and frozen -80 °C.

6.2.3 Cell wall hydrolase assay for CD2791-Lysoplate method

To assess if purified His.T7.CD2791 was correctly folded, an assay was developed based on the homology of the cell wall binding domain (found in both CD2791 and all the 28 SlpA paralogs) to N-acetylmuramoyl-L-alanine amidase. This assay is broadly the same as Osserman & Lawlor (1966) and uses the lysis of *Micrococcus lysodeikticus* cells walls immobilised in agarose.

Briefly, 50 mg *Micrococcus lysodeikticus* ATCC No. 4698 (Sigma) was re-suspended in a small volume of TK buffer (0.1 M tris-HCl, 0.1 M KCl pH 7.0). Resuspended cells were added to molten 1% (w/v) agarose in TK buffer, poured into Petri dishes, allowed to set, then 2 mm (dia.) wells were made. Each well held approximately 30 µl. Samples and controls: lysozyme (10 µg), amidase from *Pseudomonas aeruginosa* (Sigma) control (30 µg), TK buffer only control (diluted in where necessary in TK buffer) were added to the wells. Plates were incubated at room temperature for 24 hr, after which zones of α (incomplete) or β (complete) lysis (transparency) were measured.

6.2.4 Crystallisation of His.T7.rCD2791

His.T7.CD2791 was first dialysed into 50 mM HEPES pH 7.4 then concentrated to approximately 5 mg/ml, then using a nano-dispensing robot (Art Robins Instruments), sitting drop vapour diffusion crystallisation trials were set up in

96 well Intelli-Plate(s) (Art Robins Instruments). 0.15 µl (150 nl, 300 nl for 2:1 ratio) CD2791 protein solution was dispensed into each of the three wells along with an appropriate amount of reservoir solution giving 2:1, 1:1 and 1:2 protein:reservoir ratios. The following screens were initially assessed

- PACT *premier*™ (PEG/ion screen)
- JCSG-*plus*™ (sparse matrix)
- Structure Screen 1 (sparse matrix)
- Structure Screen 2 (extension of SS1 but with novel precipitants and combinations)

All of which are available from Molecular Dimensions™.

Conditions with potential results (herein referred to as hits), i.e. crystallisation conditions were close but needing to be optimised, were manually set up as hanging drop vapour diffusion crystallisations in 24-well plates. Conditions and original screen along with optimisations in x- and y- axes are given in Table 6.2.2.

6.2.5 Expression, purification and crystallisation of domains of selected surface proteins

The synthetically synthesised genes of selected surface proteins were mutated/truncated such that the predicted CWBDs were removed (CD2791 and CD2767 C-term, Cwp66- N-term) leaving only the 'functional domain'.

This was accomplished by two methods:

1. Amino acid sequences corresponding to the functional domains, flanked by NdeI and EcoRI and two stop codons, were submitted to GENEART for synthetic synthesis. The resulting gene(s) were cloned into pET28(a) using the same two restriction sites thus, providing an N-terminal His_6 tag only.

2. The same domain(s) as submitted to GENEART above were cloned from the original GENEART optimised gene(s) (Section 6.2.1.1) using primers found in

Table 6.2.3. The resulting PCR product was cloned into the pET SUMO (Invitrogen) vector via TA cloning. The resulting protein product was fused at the N-terminus to His_6 and SUMO tags.

Table 6.2.2 Manually set-up hanging drop vapour diffusion crystallisation of His.T7.rCD2791

Screen (if appropriate)	**Original condition**			**Optimisation x-axis**	**Optimisation y-axis**
	Salt	Buffer	Precipitant		
PACT *premier*™	-	0.1 M SPG buffer pH 7.0	25% w/v PEG 1500	pH 6.0, 7.0, 8.0	% PEG 1500 30, 25, 20, 15
→Manual 1	-	0.1 M SPG buffer pH 6.0	20% w/v PEG 1500	-	-
JCSG-*plus*™	-	0.1 M sodium cacodylate pH 6.5	40% v/v MPD, 5% w/v PEG 8000	pH 5.5, 6.5, 7.5	% MPD 40, 30, 20, 10
	0.2 M NaCl	0.1 M HEPES pH 7.5	10% v/v 2-propanol	pH 6.5, 7.5, 8.5	% 2-propanol 20, 15, 10, 5
	-	0.1 M tris pH 8.5	20% v/v ethanol	pH 7.5, 8.5, 9.5	% ethanol 40, 30, 20, 10
Structure Screen 1	-	-	0.2 M magnesium Formate	-	Molarity 0.4, 0.3, 0.2, 0.1
→ Manual 2	-	-	0.2 M sodium formate	-	Molarity 0.4, 0.3, 0.2, 0.1
→ Manual 3	-	-	0.2 M magnesium sulphate	-	Molarity 0.4, 0.3, 0.2, 0.1

Table 6.2.3 Primers used to remove selected surface proteins CWBD and for insertion into pET SUMO expression vector

Gene	Predicted CWBDs[a]	Region Cloned	Primers for pET SUMO TA cloning[b]	Approx MW [amino acids]
cd2791	326-417 426-512	27-322	2791_SUMO_F AGC ACC ACC CAG GTG AAA AAA GAA A 2791_SUMO_R TTA TTT GCT GTT GCC TTC CAG CG	32.0 kDa [298]
cwp66	33-121 130-221 228- 322	324-585	*cwp66*_SUMO_F AGC ATT GAT ATG CAG GAA GA *cwp66*_SUMO_R TTA GAT GTG ATC CAT TTT CGG	29.9 kDa [264]
cd2767	402-493 505-596 609-684	27-401	2767_SUMO_F AGC AAC GAT AAA GAA ATG CG 2767_SUMO_R TTA CAC TTT CAC GTT GCT CG	42.5 kDa [376]

[a] as given by *C. difficile* 630 genome sequence (NCBI)

[b] with addition of N-terminal Serine (AGC) to improve SUMO protease cleavage

6.2.5.1 Expression of Purification of domains of selected surface proteins

6.2.5.1.1 His tagged only (pET28a) constructs

Expression was as described in General Expression protocol (Section 6.2.2.1). IMAC purification of functional domains was performed as described in General Methods Section 2.4.1.3. His.rCD2767$_{27\text{-}401}$ was also purified as described (Section 2.4.1.3) but using 50 mM sodium phosphate (monobasic) pH 7.5 in place of 50 mM tris pH 8.0, as tris is a known inhibitor of alpha amylase activity.

6.2.5.1.2 pET SUMO (His.SUMO) constructs

Expression was as described in General Expression protocol (Section 6.2.2.1) except using *E. coli* BL21 (DE3). His.SUMO tagged proteins were first purified by IMAC as

per General Methods Section 2.4.1.3. Collected eluted fractions were then dialysed into SUMO protease cleavage buffer (50 mM tris-HCl, 150 mM NaCl pH 8.0, 0.2% igepal, 1 mM DTT). A second round of IMAC was performed using 50 mM tris-HCl, 150 mM NaCl pH 8.0 as binding/wash buffer to ensure any uncleaved material and/or contaminants bound, with cleaved untagged protein not binding. To remove residual detergent from the cleavage buffer, the flow through was dialysed into IEX (Q pH 8.0) start buffer and purification undertaken as per General Methods (Section 2.4.2).

6.2.5.2 Crystallisation of domains of selected surface proteins

6.2.5.2.1 Concentration

6.2.5.2.1.1 rCD2791$_{27\text{-}322}$

Post Q purification, His.SUMO derived rCD2791$_{27\text{-}322}$ was concentrated in a Vivaspin 20 10k MWCO spin concentrator then dialysed into 20 mM tris pH 8.0 (final concentration 12.04 mg/ml (A_{280}) (or ~5mg/ml using ProtParam (Gasteiger *et al.* 2005) calculated extinction co-efficient 0.419) before sitting drop vapour diffusion crystallisation as detailed in Section 6.2.5.3.

6.2.5.2.1.2 His.rCD2767$_{27\text{-}401}$

Post IMAC, His.rCD2767$_{27\text{-}401}$ was dialysed into 50 mM tris 150 mM NaCl pH 8.0 (or 50 mM MES 150 mM NaCl pH 6.0) 0.2 µm filtered then concentrated in a Vivaspin 20 10 kDa MWCO spin concentrator (Table 6.2.4)

Table 6.2.4 Assessment of concentration of His.rCD2767$_{27\text{-}401}$

Pre concentration				
Dilution	A280	Approx mg/ml*	Volume	Total protein (mg)
1:10	0.857	3.87	20 ml	171.4 mg
Post concentration				
1:100	1.678	75.7	750 µl	125 mg
1:200	1.060	95.76	750 µl	159 mg
1:500	0.533	120.2	750 µl	199 mg

* using ProtParam (Gasteiger *et al.* 2005) extinction co-efficient of 2.216

6.2.5.3 Crystallisation set-up

Using a nano-dispensing robot (Art Robins Instruments), sitting drop vapour diffusion crystallisation trials were set up in 96 well Intelli-Plate(s) (Art Robins Instruments). 0.15 µl (150 nl, 300 nl for 2:1 ratio) protein solution was dispensed into each of the three wells along with an appropriate amount of reservoir solution giving 2:1, 1:1 and 1:2 protein:reservoir ratios. The following screens were assessed:

- PACT *premier*™ (PEG/ion screen)
- JCSG-*plus*™ (sparse matrix)
- Structure Screen 1 (sparse matrix)
- Structure Screen 2 (extension of SS1 but with novel precipitants and combinations)
- Memgold (sparse matrix membrane protein specific screen covering a range of pH, PEGs and salt additives)
- Morpheus™ (3D grid design covering a range of pH, PEGs and salt additives)

All of which are available from Molecular Dimensions™.

Conditions with potential results (hits) i.e. crystallisation conditions were close but needing to be optimised, were set up as hanging drop vapour diffusion crystallisations manually in 24-well plates (Table 6.2.5).

6.2.5.4 His.rCD2767$_{27\text{-}401}$ single crystal data collection and processing

A total of 250 images from a single crystal of His.rCD2767$_{27\text{-}401}$ (in 50 mM tris 150 mM NaCl pH 8.0 from well D10 (Structure Screen 1 & 2, 0.05 M potassium dihydrogen phosphate, 20 % w/v PEG 8000)) were recorded to a resolution of approximately 2 Å, with an oscillation angle of 1° per image on the ϕ axis (perpendicular to the x-ray beam), with a crystal to detector distance of 300.51 mm and exposure time of 2.6 sec, at 100K on beamline IO4-1 at the Diamond Light Source (Didcot, Oxon).

Diffraction images were processed (indexing, cell-refinement and integration) with iMOSFLM v1.0.4 (Leslie 2006) in each of the proposed space groups (Table 6.3.5).

Data were scaled in all possible space groups using SCALA (CCP4 1994) and taken for further analysis including molecular replacement and cell content analysis. Preliminary molecular replacement trials were performed using the PHENIX package of crystallography programs (Adams *et al.* 2010).

Table 6.2.5 Manually set up vapour diffusion crystallisation of rCD2791$_{27-322}$

Original Screen well	**Buffer/Salt**	**Buffer/Salt**	**pH**	**Precipitant**	**Protein:Reservior**
PACT Premier B3	0.1 M MIB* buffer	None	6.0	25-10% w/v PEG1500	1:2
PACT Premier C3	0.1 M PCTP* buffer	None	6.0	25-10% w/v PEG1500	1:2, 1:3
PACT Premier D3	0.1 M MMT* buffer	None	6.0	25-10% w/v PEG1500	1:2, 1:3
→Manual 1	None	None	-	25-10% w/v PEG1500	1:2

*MIB - malonic acid, imidazole, boric acid, PCTP - sodium proponate, sodium cacodylate, bis-tris propane, MMT- malic acid, MES, tris.

6.3 Results

The following CWBD containing surface proteins were chosen for crystallisation, based on their likelihood to be implicated in adherence or pathogenesis and/or immuno-stimulatory ability:

- CD2791 (66 kDa)- an uncharacterised immunogenic surface protein
- Cwp66 (66 kDa)- a heat shock inducible adhesin
- CD2767 (77 kDa)- an uncharacterised putative glycoside hydrolase

Two tracks were used to express and purify protein(s) for crystallisation: Expression of the full length protein (minus signal peptide) containing CWBDs, which may potentially yield structural information about both the functional domain and vital data regarding the CWBD(s) and expression of only the 'functional' domain (without the CWBD (and signal peptide)).

6.3.1 CD2791

6.3.1.1 Expression and purification of rCD2791 (full length)

rCD2791 expressed as a clear band at 66 kDa when fused to N-terminal His_6 and T7 tags (His.T7.rCD2791). IMAC purified His.T7.rCD2791 plus truncated material between 28 kDa - 66 kDa which are detected with anti-His, anti-T7®Tag and anti-R20291 SLP IgG in western blot (Figure 6.3.1). To remove degraded/truncated species two IEX chromatography (Q-sepharose or SP-Sepharose) resins were tested, however neither satisfactorily isolated pure His.T7.rCD2791, in both cases His.T7.rCD2791 eluted as a single peak with contaminating material.

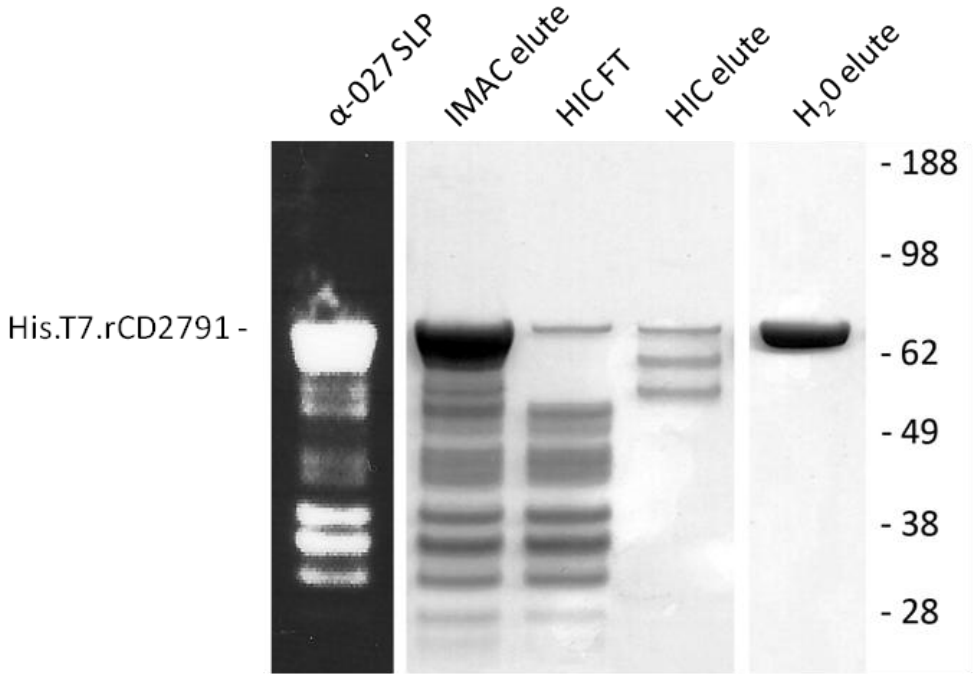


Figure 6.3.1 His.T7.rCD2791 purification. Lane 1 Western blot with anti-R20291 SLP IgG of Lane 2. Lane 2 IMAC elute showing extensive N-terminal degradation/truncation, Lane 3 and 4 material not bound to or eluted during HIC purification, Lane 4 Water eluted 66 kDa (His.T7.rCD2791) material from the same phenyl sepharose HP (HIC) column.

Using hydrophobic interaction chromatography a fraction of His.T7.rCD2791 appeared to bind so strongly it could only be eluted with water, this lead to >90% His.T7.rCD2791 purity by SDS-PAGE (Figure 6.3.1). Contaminating material either did not bind or eluted during the desalting gradient.

6.3.1.1.1.1 Assay for N-acetylmuramoyl-L-alanine amidase activity

It has been shown previously that the HMW SLP (which consists of three CWBDs) exhibits N-acetylmuramoyl-L-alanine amidase (amidase) activity in zymogram assay against *Micrococcus luteus* cells (Calabi *et al.* 2001). The amidase activity of the two C-terminal CWBDs of CD2791 were therefore available to help determine if the His.T7.rCD2791 purified was correctly folded.

A method using immobilised *Micrococcus lysodeikticus* (also known as *Micrococcus luteus*) was 'the lysoplate' developed, based on serum and urinary lysozyme (muramidase) activity (Osserman & Lawlor 1966), where cell lysis/amidase activity results in transparent area(s) in a solid phase agarose- *Micrococcus* cell gel matrix (Table 6.3.1).

Table 6.3.1 Lysoplate results for rCD2791 fractions

Well number	α-lysis (incomplete) (mm diameter)	β-lysis (complete) (mm diameter)
1- 10 µg Lysozyme control	39	22
2- 30 µg Amidase* control	39	1
3- 11.19 µg water eluted CD2791	13	0
4- 31.08 µg water eluted CD2791	14	0
5- 10.75 µg flow-through material	0	0
6- Buffer only	0	0

* Amidase from *Pseudomonas aeruginosa* (Sigma- A6691). TK buffer 0.1 M tris-HCl, 0.1 M KCl pH 7.0 (Fan & Beckman 1973).

After 24 hr, water eluted 66 kDa (His.T7.rCD2791) gave a zone of lysis approximately one third that of the amidase control (at both concentrations tested). The HIC non-binding or 'normally' eluted material did not show any lytic activity. Varying pH (6, 7 or 8) and/or including 1 mM $ZnCl_2$ in the reaction buffers did not alter lysis zones for His.T7.rCD2791 (zinc or other divalent cations are a known requirement of N-acetylmuramoyl-L-alanine amidases (Shida *et al.* 2001; Kerff *et al.* 2010)). This data suggests that the two C-terminal CWBDs, of HIC water eluted His.T7. CD2791, were at least partially correctly folded and able to lyse *Micrococcus* cell walls in a manner analogous to the HMW SLP. Consequently, His.T7.rCD2791 was dialysed into 10 mM HEPES pH 7.4, concentrated to approximately 6 mg/ml and put forward into crystallisation.

6.3.1.1.1.2 Crystallisation of rCD2791

Robotic assisted sitting drop crystallisation trials (570 conditions in total over 5 commercially available crystallisation screens) revealed a range of buffer conditions

(~20) where crystallisation may occur. Hits were primarily sea urchin like or fine tiny needles (Table 6.3.2).

There did not appear to be a consistent feature of the hits, e.g. buffer or precipitant, other than a pH requirement between pH 6.5 - 8.5. Although, approximately one third of the hits included magnesium chloride and/or tris based buffers. Manually set up hanging drop crystallisation, optimising buffer composition, pH and precipitant concentration did not change crystal quality or size and yielded small needles (See Figure 6.3.2).

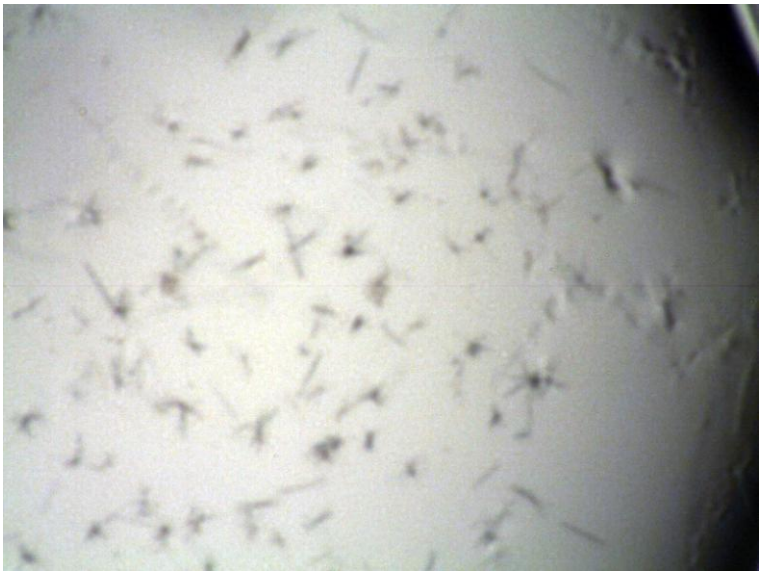


Figure 6.3.2 His.T7.rCD2791 needle crystals. Obtained from manually set-up hanging drop crystallisation trial with His.T7.rCD2791 (6.02 mg/ml in 10 mM HEPES pH 7.0) in 0.2 M NaCl, 0.1 M HEPES, 10% isopropanol pH 7.5 (derived from JCSG-plus™ sparse matrix screen).

6.3.1.1.1.3 Optimisation of rCD2791 for crystallisation

Due to the limited crystallisation displayed in sitting drop crystallisation screens it was suspected that crystallisation was being hindered by impurities in the protein sample or intra/inter protein or a combination of both. Therefore, methods were used to assess the native state of the protein and to improve the purity.

Table 6.3.2 Sitting drop crystallisation conditions which demonstrated varying degrees of crystallisation of His.T7.rCD2791

		Reservoir conditions				
Screen	**Well**	***Salt***	***Buffer***	***pH***	***Precipitant***	**Result**
Structure Screen 1 & 2	D12	-	-		0.2 M magnesium formate	Needles
	E4	0.2 M magnesium chloride	0.1 M tris	8.4	3.4 M 1,6-hexanediol	Clusters
	E9	-	0.1 M tris	8.5	20% ethanol	Clusters
	F7	-	0.1 M HEPES	7.5	10% PEG 8000, 8% ethylene glycol	One lone needle
JCSG-plus™	B10	0.2 M magnesium chloride	0.1 M sodium cacodylate	6.5	50% PEG200	Needles
	B8	0.2 M magnesium chloride	0.1 M tris	7	10% PEG8000	Needles
	C8	-	0.1 M tris	8.5	20% v/v ethanol	small cluster of needles
	E3	0.2 M NaCl	0.1 M HEPES	7.5	10% v/v 2-propanol	Needle clusters
	E9	-	0.1 M MES pH 6.5	6.5	1.6 M magnesium sulphate	Needles
	F6	-	0.1 M bicine pH 9.0	9	10% MPD	sea urchins
	F5	0.2 M magnesium chloride	0.1 M tris pH 8.5	8.5	50% ethylene glycol	crystaloids
PACT premier™	G12	-	0.1 M SPG buffer	7	25 % w/v PEG 1500, 25 % w/v PEG 1500	Sea Urchins
	A4	0.2 M sodium malonate	0.1 M bis tris propane	7.5	20 % w/v PEG 3350	Tiny rods

6.3.1.1.1.3.1 Size exclusion chromatography (SEC)

To assess if the protein sample was aggregated, SEC was used. SEC of 'pure' HIC water eluted His.T7.rCD2791 revealed that the protein was predominantly aggregated and came out in the void volume. The second largest species (50% peak size) correlated with a size of a dimeric species 163 kDa (132 kDa predicted) followed by a small amount of monomeric species (74.8 kDa, 66 kDa predicted) and smaller impurities (Figure 6.3.3).

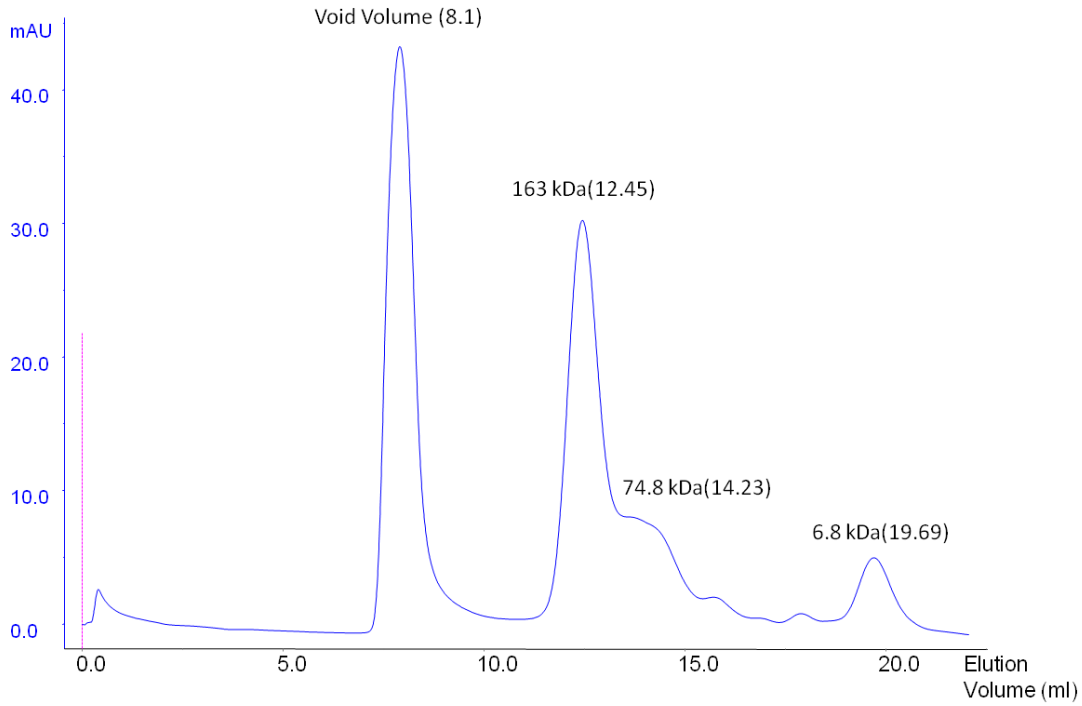


Figure 6.3.3 Size exclusion chromatography (SEC) of His.T7.rCD2791. His.T7.rCD2791 in 10 mM HEPES pH 7.4 (6 mg/ml) using a Superdex 200 10/300 GL column equilibrated with PBS. Numbers in brackets represent elution volume (ml).

Therefore, the presence of aggregated material was likely to have affected both the ability of His.T7.rCD2791 to crystallise, i.e. pack into a regular periodic structure, and possibly affected the amidase activity, due to a lack of free-unaggregated material to cleave *Micrococcus* cell walls.

6.3.1.1.1.3.2 IMAC optimisation

The hydrophobic nature of His.T7.rCD2791 observed during HIC purification and the aggregation seen in SEC, was suspected to be due to the CWBDs aggregating with

themselves or other proteins. Certain surface proteins containing CWBDs, e.g. SlpA, CD2791 or CD2784, can be removed from the *C. difficile* cell surface using low pH or chaotropic agents, i.e. disrupt CWBD interactions, a range of protease inhibitors, sugars, detergents, kosmotropic and chaotropic additives (as suggested by Bondos & Bicknell (2003)), were added to cell lysis and IMAC buffers to potentially decrease aggregation and degradation. A single 66 kDa species was obtainable, using all additives, as a secondary and tertiary IMAC peak(s) when using gradient elution on small (50 ml) culture volumes, with 1% CHAPS giving the largest (mAU) yield (Figure 6.3.4A).

Upon scale up (1litre), the addition of 1% CHAPS did not appear to result in the same triphasic elution and His.T7.rCD2791 eluted impurely (Figure 6.3.4B). Further purification with high resolution MonoQ, with buffers also supplemented with 1% CHAPS, did not purify His.T7.rCD2791 away from contaminants.

Denaturing purification and on-column refolding were assessed twice. The first run eluted microgram amounts of pure His.T7.rCD2791, not enough for crystallisation. During the second run, after removing denaturant, His.T7.rCD2791 and contaminants eluted from the column before elution with imidazole.

Full length, C-terminally His_6 tagged rCD2791, may have expressed and purified without the N-terminal truncates seen for His.T7.rCD2791, interactions of the CWBDs were still likely to have impeded obtaining pure (un-aggregated) protein.

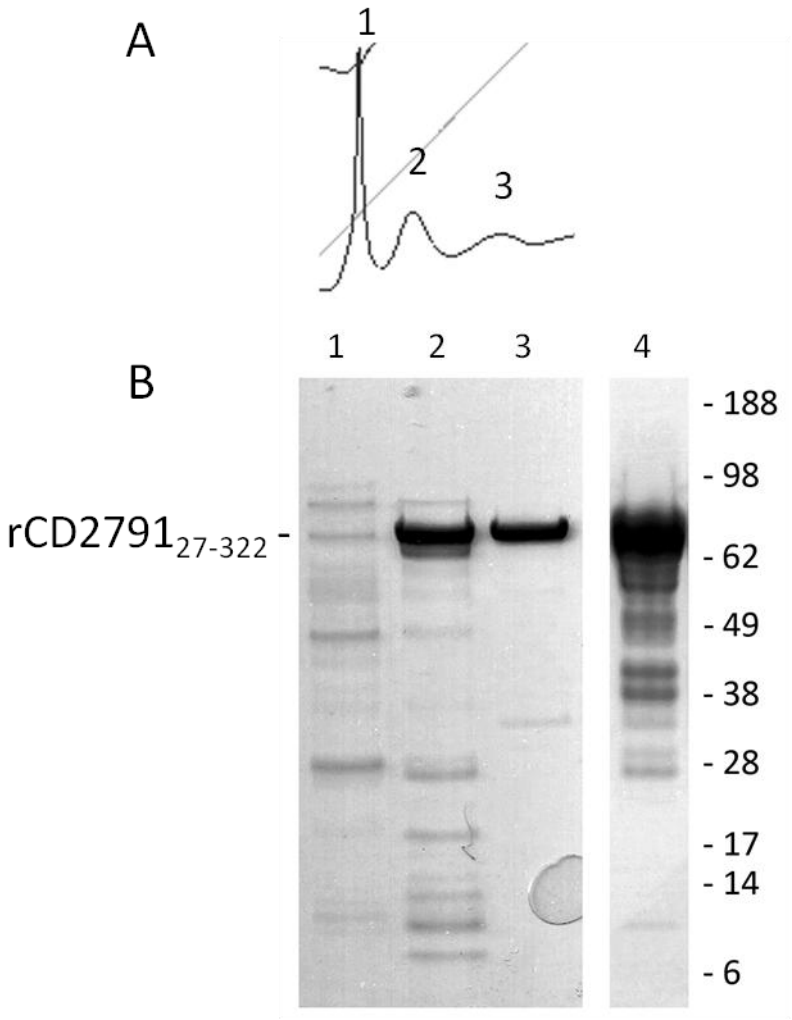


Figure 6.3.4 IMAC purification of His.T7.rCD2791 with buffers supplemented with 1% CHAPS. (A) Chromatograph showing elution peaks (B) SDS-PAGE analysis of fractions from small scale test (50 ml culture, 1 ml HisTrap HP) (Lane 1-3) large scale purification (1 litre culture, 5 ml HisTrap) (Lane 4).

6.3.1.1.2 Expression and purification of the N-terminal domain of CD2791 (rCD2791$_{27-322}$)

Due to the difficulties in obtaining pure non-aggregated full length rCD2791, expression of only the 'functional' domain was assessed, i.e. minus the CWBDs, from residue T27-K322 (rCD2791$_{27-322}$). T27 represents the predicted start of CD2791 after the signal peptide, although A25 has been shown to be the actual starting N-terminal residue (Calabi & Fairweather 2002).

A His_6 tagged version (His.rCD2791$_{27-322}$) expressed and purified as a 45 kDa species (~33 kDa predicted) in IMAC with a large amount of truncated (<30 kDa) species. Dialysis into an IEX (MonoQ pH 8.0) buffer appeared to precipitate both contaminating species and a significant proportion of His.rCD2791$_{27-322}$. The remaining His.rCD2791$_{27-322}$ eluted as a single peak with a 1 M NaCl gradient giving a prominent band at 45 kDa with contaminating species at 35, 30, 25 and 20 kDa in SDS-PAGE (Figure 6.3.5).

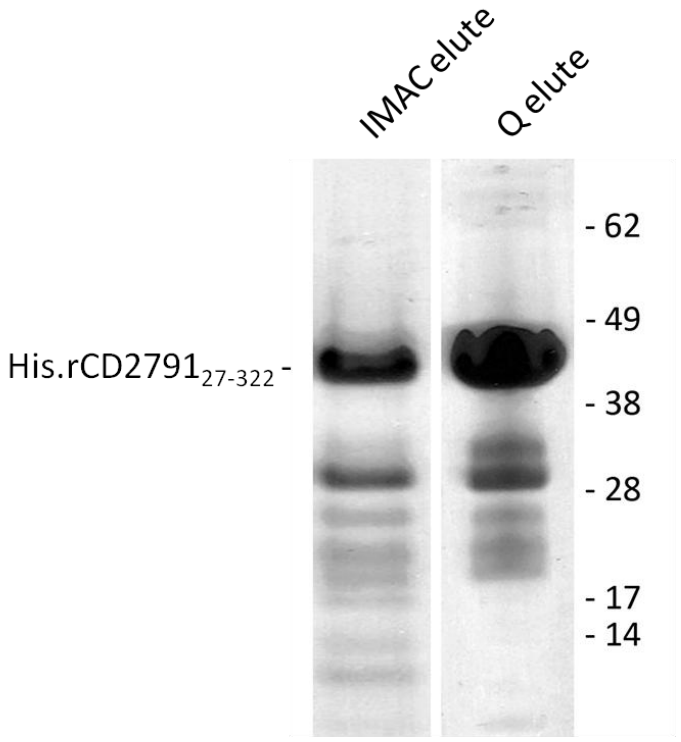


Figure 6.3.5 Purification of His.rCD2791$_{27-322}$. Lane 1- Material eluted from IMAC. Lane 2- Unprecipitated material eluted from Mono Q.

To prevent the precipitation seen post IMAC and to try to decrease N-terminal truncation, fusion to a small ubiquitin-like modifier (SUMO) tag was assessed, as this protein has been shown to aid purification of difficult proteins (Butt *et al.* 2005; Marblestone *et al.* 2006). His.SUMO.rCD2791$_{27-322}$ expressed well and purified as a 52 kDa species in IMAC. Subsequent removal of the His.SUMO tag with SUMO protease and a second round of IMAC (cleavage removes the His_6 tag so rCD2791$_{27-322}$ flows through) left a 40 kDa species (predicted 32 kDa) with 84, 30, 28, 25 and 20 kDa impurities. IEX purification to remove SUMO protease buffer, eluted a single

peak containing a main 40 kDa band (rCD2791$_{27-322}$) plus other minor species at approximately 29 and 31 kDa (Figure 6.3.6). The ratio of 40 kDa rCD2791$_{27-322}$ to other species decreased across the eluted peak. After dialysis into 10 mM tris pH 8.0 and concentration to 12.04 mg/ml, the eluted rCD2791$_{27-322}$ material was used in crystallisation trials. Interestingly, the final His.SUMO version of rCD2791$_{27-322}$ demonstrates a similar SDS-PAGE profile to the MonoQ purified (un-precipitated) His.rCD2791$_{27-322}$ purified in Figure 6.3.5.

SEC of His.SUMO.rCD2791$_{27-322}$ and the native (post tag removal) rCD2791$_{27-322}$ gave monomeric species of approximately 74 kDa and 44 kDa respectively, putatively supporting SDS-PAGE results suggesting that rCD2791$_{27-322}$ runs larger than the predicted molecular weight (32 kDa).

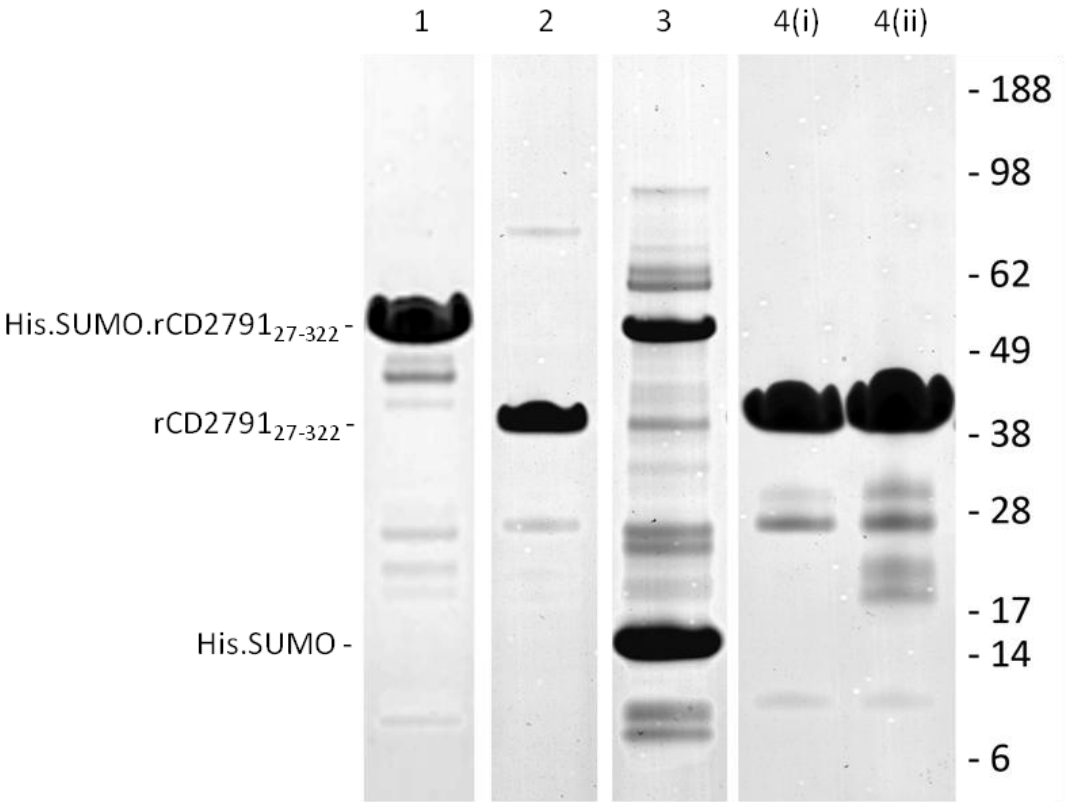


Figure 6.3.6 Purification of His.SUMO tagged rCD2791$_{27-322}$. Lane 1- IMAC elute, Lane 2- rCD2791$_{27-322}$ Post His.SUMO cleavage IMAC flow through, Lane 3- Post His.SUMO cleavage IMAC elute showing uncleaved full length and liberated His.SUMO. Lane 4(i & ii)- IEX (Q, pH 8.0) purification of rCD2791$_{27-322}$ IMAC flow through.

6.3.1.1.2.1 Crystallisation of rCD2791$_{27\text{-}322}$

Robot setup sitting drop crystallisation screens of rCD2791$_{27\text{-}322}$ appeared to have a number of hits (Table 6.3.3), particularly in proprietary buffers from the PACT premier screen (Figure 6.3.7). However, manual setup and optimisation of these conditions all resulted in protein precipitation.

Table 6.3.3 Sitting drop crystallisation conditions which demonstrated varying degrees of crystallisation rCD2791$_{27-322}$

Reservoir conditions						
Screen	**Well**	***Salt***	***Buffer***	***pH***	***Precipitant***	**Result**
JCSG+	F9	None	None	7.0	2.4 M sodium malonate	Needles
	H9	0.2 M lithium sulphate	0.1 M bis tris	5.5	25 % w/v PEG 3350	Needles
	H10	0.2 M ammonium acetate	0.1 M bis tris	5.5	25 % w/v PEG 3350	Needles
Memgold	A3	None	0.015 M tricine	8.5	24 % w/v PEG 4000	Needles
	C7	None	0.02 M bis tris	7.0	15 % w/v PEG 2000	'Sea Urchins'
	D3	None	0.05 M HEPES	7.5	22 % v/v PEG 4000	Needles
	F8	0.1 M sodium chloride/ 0.005 M magnesium chloride	0.1 M tris	8.5	30 % w/v PEG 2000 MME	Needles
PACT premier	B3-5	0.1 M MIB* buffer	None	7.0-9.0	25% w/v PEG1500	Figure 3.3.4
	C3-5	0.1 M PCTP* buffer	None	7.0-9.0	25% w/v PEG1500	Figure 3.3.4
	D3-5	0.1 M MMT* buffer	None	7.0-9.0	25% w/v PEG1500	Figure 3.3.4
Screen	**Well**	***Ligand stock***	***Buffer stock***	***pH***	***Precipitant stock***	**Result**
Morpheus[a]	H5	0.1 M amino acids	0.1 M Buffer 2	7.5	30% P550MME_P20K	Needles with spheroids
	G1	0.1 M carboxylic acids	0.1 M Buffer 1	6.5	30% P550MME_P20K	Spheroids
	H9	0.1 M amino acids	Buffer 3	8.5	30% P550MME_P20K	Needles

*MIB - malonic acid, imidazole, boric acid, PCTP - sodium proponate, sodium cacodylate, bis-tris propane, MMT- malic acid, MES, tris.

[a] For details of the composition of the buffers in the Morpheus screen please see Molecular dimensions screen datasheet.

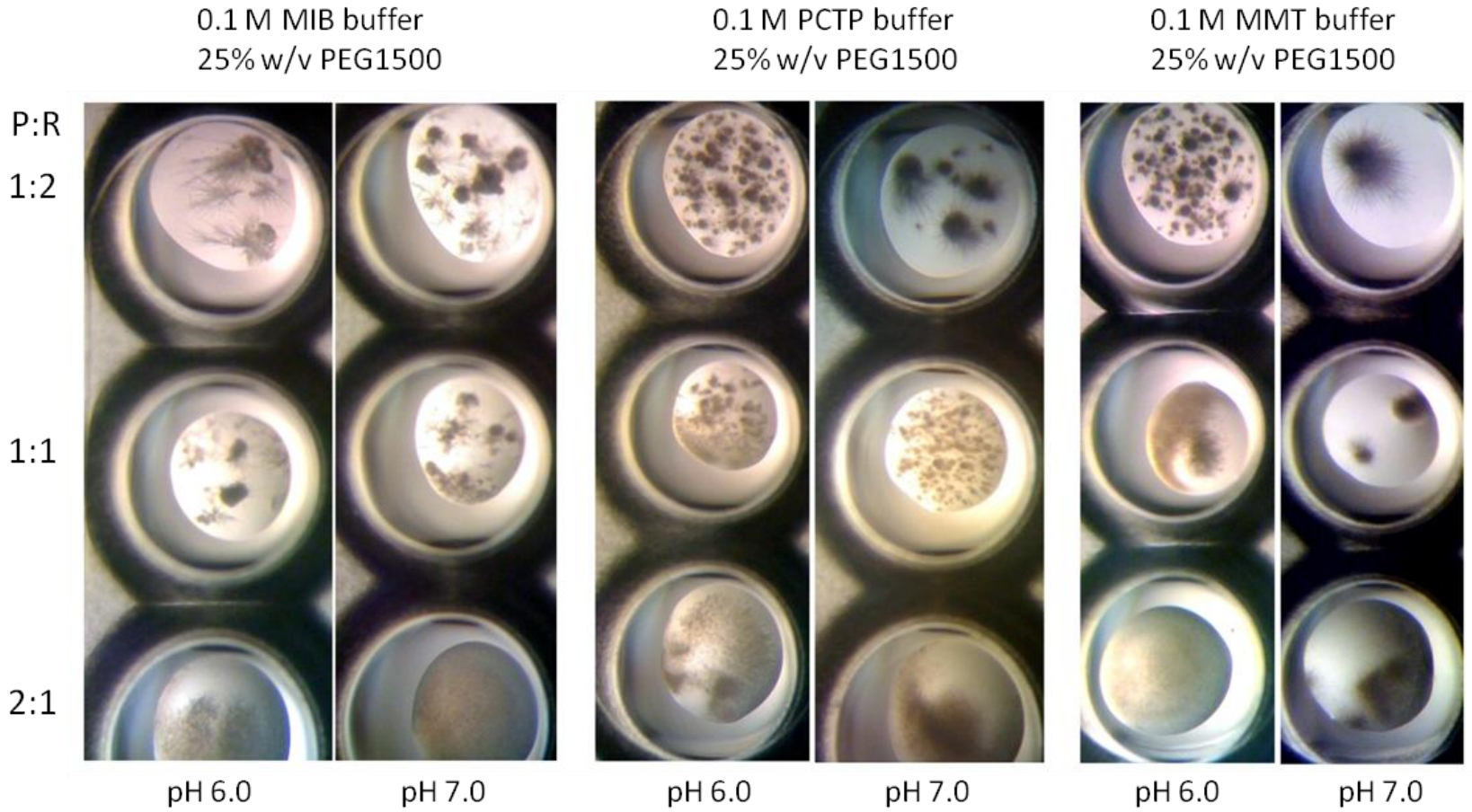


Figure 6.3.7 Sitting drop crystallisation of rCD2791$_{27\text{-}322}$ using the PACT premier™ PEG/Ion screen. Protein (P) : Reservoir solution (R) ratio.

6.3.1.1.3 Modelling of CD2791$_{27\text{-}322}$

Previous BLASTP analysis of the CD2791 N-terminal region revealed homology (39% amino acid similarity) to both the 125 kDa SLP of *Bacillus sphaericus* and the transducer of rhodopsin (Htr) II from the archaebacterium *Natronobacterium pharaonis* (Calabi & Fairweather 2002).

BLASTP, in this chapter, suggests that the C-terminal region of CD2791$_{27\text{-}322}$ (Y124-E318) shares approximately 30% identity with the S-layer (Pfam 04122 containing) proteins of *Clostridium hiranonis* DSM 13275 and a ~50 aa region (between V221-S270) shares approximately 40% identity with the cytotoxin associated protein CagA of *Helicobacter pylori* (Backert *et al.* 2010). CD2791$_{27\text{-}322}$ shares approximately 17% identity to the LMW SLP. The full length (minus signal peptide) CD2791 and SlpA share 29.4% identity, although this is largely through the CWBDs.

Performing a BLAST search of the sequences derived from structural domains in CATH (Protein structure classification database, www.cathdb.info) reveals that CD2791$_{27\text{-}322}$ has a 41 amino acid region with 36.59% identity with both 1KIX (Peersen *et al.* 2002) and 1JB7 (Horvath & Schultz 2001), both of which are telomere binding proteins from the ciliate protozoa *Oxytricha nova*.

Therefore, BLAST analysis suggests that CD2791$_{27\text{-}322}$ shares little homology with known proteins at sequence level. Using tertiary structure homology modelling programs to try to reveal protein similarities confounds the lack of similar proteins as SWISS-MODEL (Bordoli *et al.* 2009) and ESyPred3D (Lambert *et al.* 2002) could not identify useful template structures. CPHmodels (Nielsen *et al.* 2010), using remote homology modelling, suggests a poor quality model (Z-score 5.6, Z-score 10+ indicates a high reliability model) based on *Listeria* invasion protein InlB (2UZX) (Figure 6.3.8).

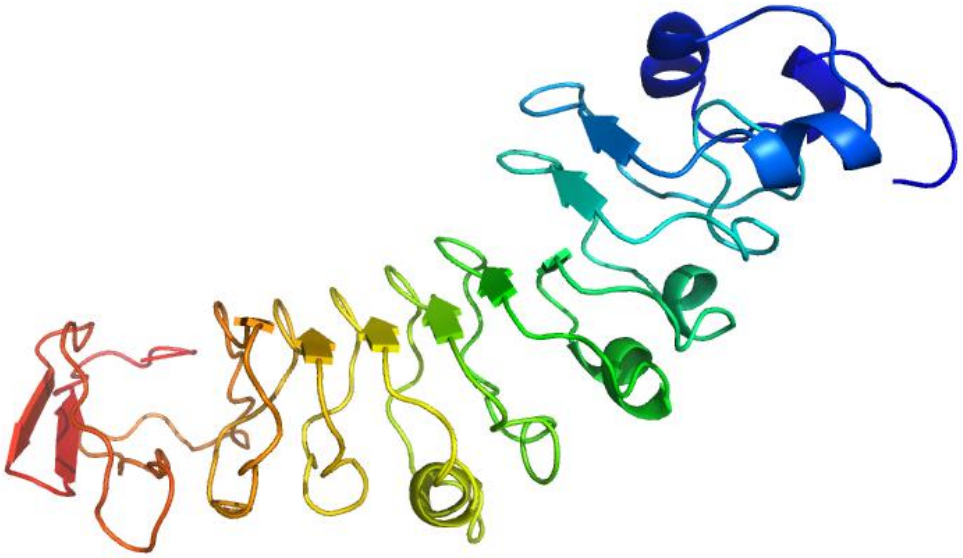


Figure 6.3.8 Predicted structure of CD2791$_{27\text{-}322,}$ homology modelled on Listeria invasion protein InlB (2UZX) (by CPHmodels-3.0)(Nielsen *et al.* 2010). Prepared using PyMOL (www.pymol.org) (N- blue C- red).

Threading (fold recognition) servers also found little to no similarity to known folds, as found with homology searches. Both Phyre (Kelley & Sternberg 2009) and HHPred (Soding 2005) found small regions with limited fold similarity to known proteins. However, mGenThreader (Lobley *et al.* 2009) which uses fold recognition and identification of distant homologues using profile-profile alignments and predicted secondary structure (using PSIPRED (Jones 1999)), suggests that the secondary structure of CD2791$_{27\text{-}322}$ is similar to the S-layer protein SbsC of *Geobacillus stearothermophilus* (2RA1), sufficiently that the fold match is rated 'High' (p value <0.001) (while sharing 10.5% identity) (Figure 6.3.10). Entering the alignment into MODELLER (Eswar *et al.* 2007) and aligning it with SbsC (Pavkov *et al.* 2008), reveals CD2791$_{27\text{-}322}$ aligns with domains II and III of rSbsC$_{(31–443)}$, although the N-terminus of CD2791$_{27\text{-}322}$ (see secondary structure prediction Figure 6.3.11) aligns with the alpha-helical SCWP binding domain I of SbsC (Figure 6.3.9).

Analysis of just the secondary structure prediction (Figure 6.3.11) (PSIPRED (Jones 1999)) suggests CD2791$_{27\text{-}322}$ has two sheet-helix-sheet repeats at the N-terminus

then a predominantly β -sheet region ending with a helix, which putatively marks the start of the largely α-helical CWBDs. Secondary structure prediction suggests, that after the signal peptide has been removed, the N-terminus of CD2791 contains a helix and loop 'head' domain followed by a stem β sheet domain, then the two CWBD. It is interesting to that the both the predicted secondary structure of $CD2791_{27\text{-}322}$ and the LMW SLP (Fagan *et al.* 2009) are both two domain proteins, suggesting that this type of arrangement may be involved in S-layer array packing.

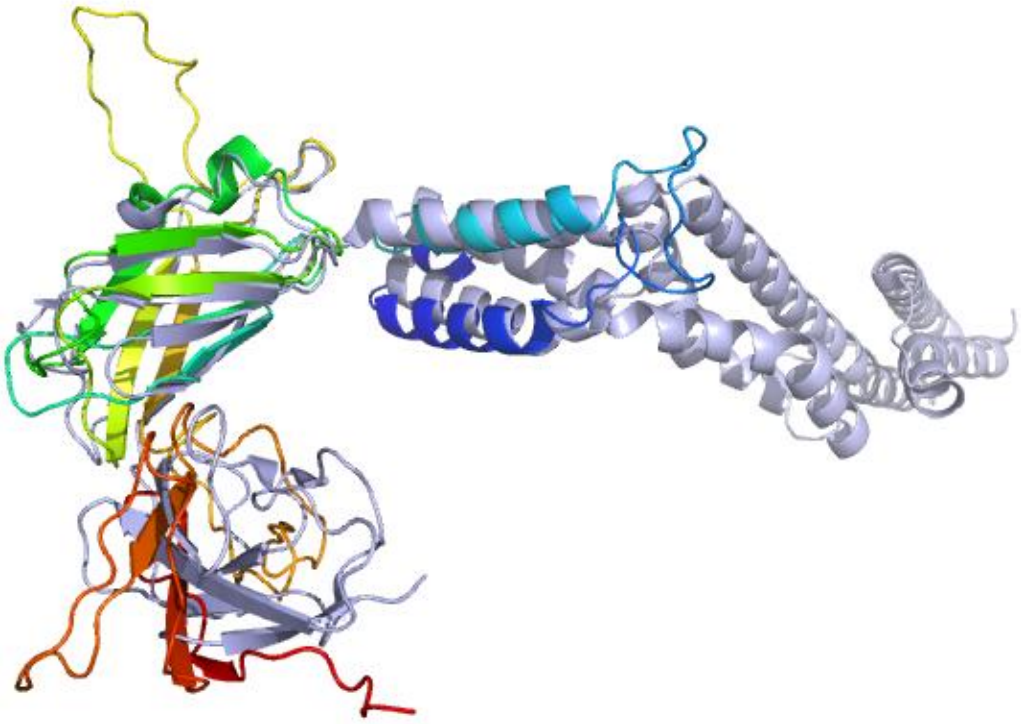


Figure 6.3.9 Superposition of pGENThreader (Lobley *et al.* 2009) secondary structure fold alignment of $CD2791_{27\text{-}322}$ (Blue N- Red C-term, modelled with MODELLER (Eswar *et al.* 2007)) with S-layer protein SbsC of *Geobacillus stearothermophilus* (2RA1) (Grey). Prepared using PyMOL (www.pymol.org) (N- blue C- red).

The lack of reference structures upon which CD2791 can be modelled (by either homology or fold modelling), highlights the need for a crystal structure to enable novel fold identification or allow understanding of function based on conserved folds not identifiable by current predictive modelling servers.

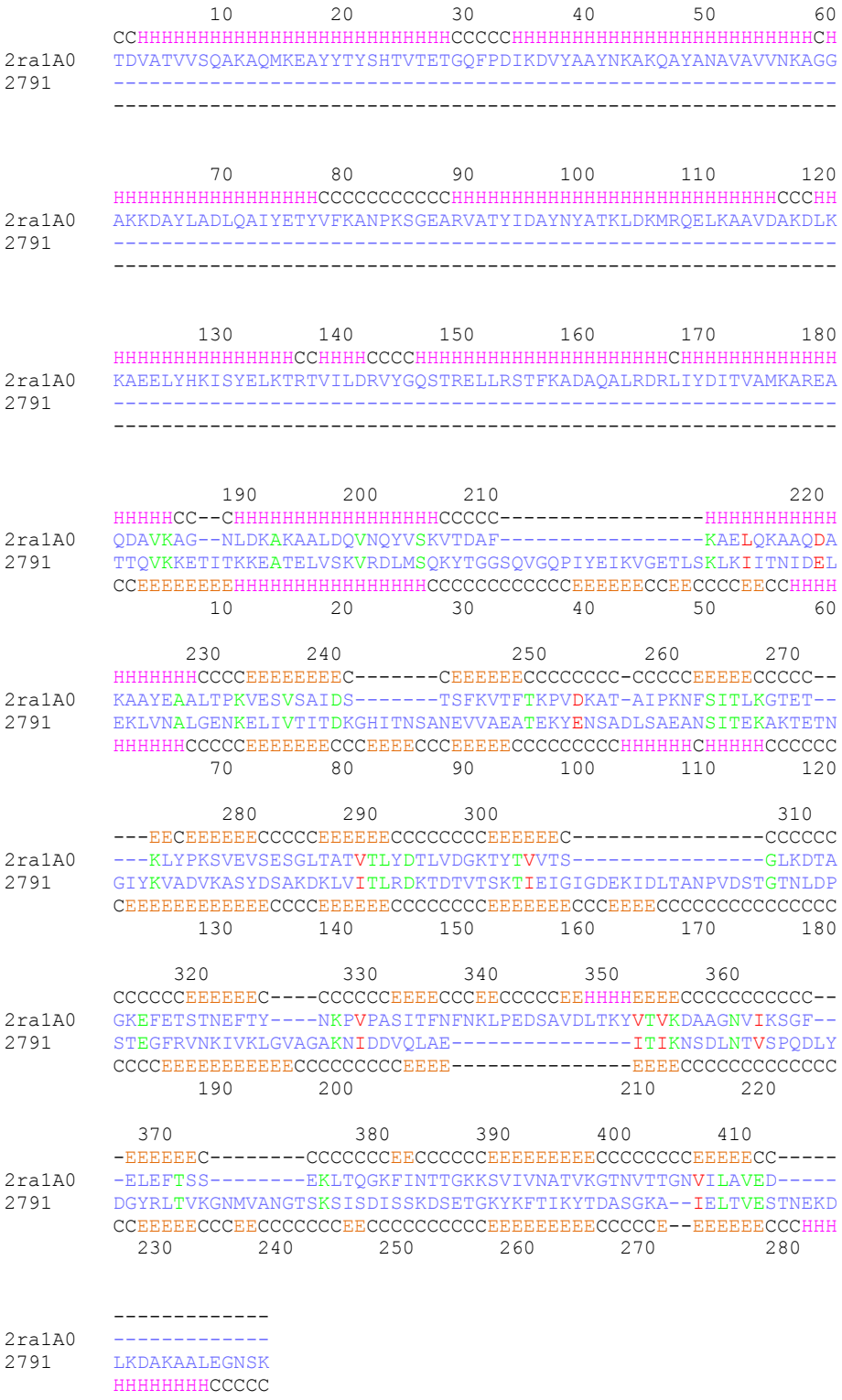


Figure 6.3.10 pGENThreader (Lobley *et al.* 2009) alignment of $CD2791_{27\text{-}322}$ with the S-layer protein SbsC of *Geobacillus stearothermophilus* (2RA1) (Pavkov *et al.* 2008).

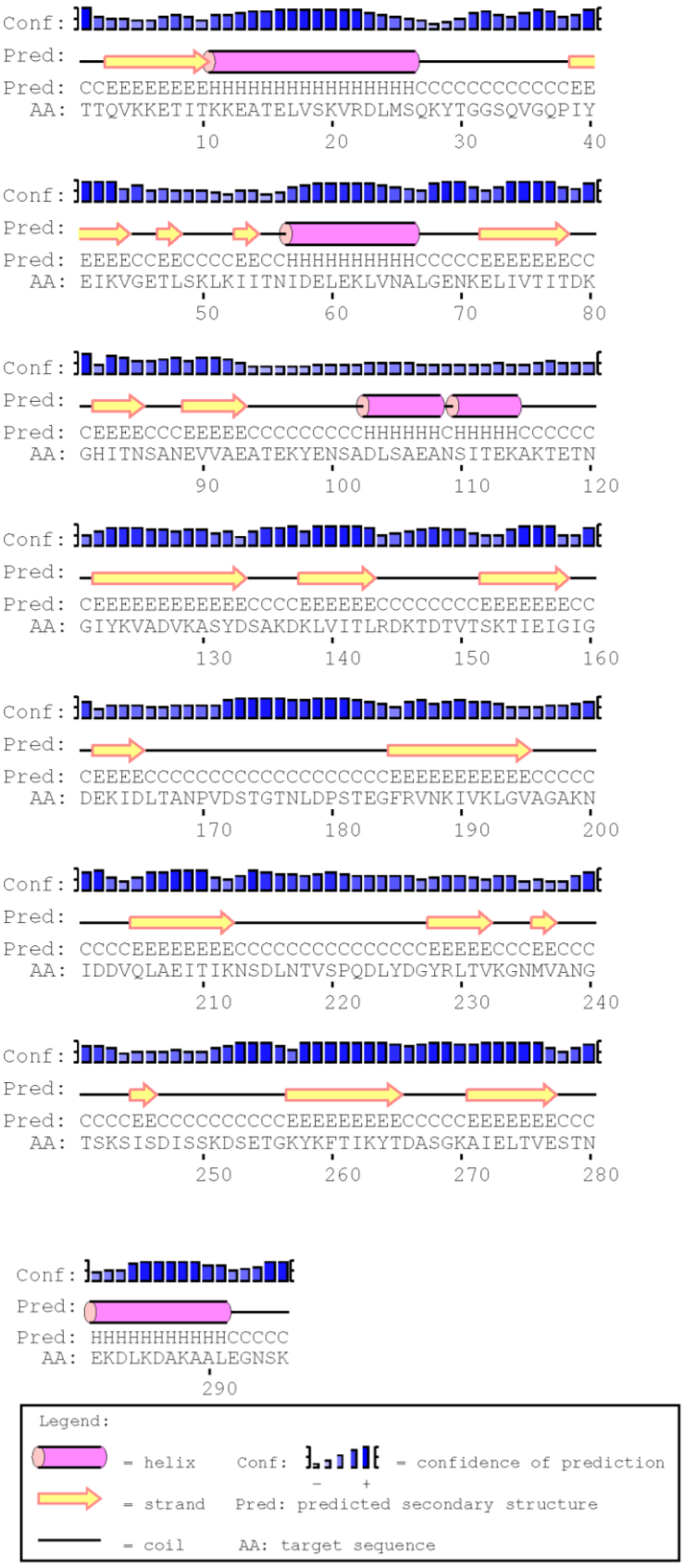


Figure 6.3.11 Secondary structure prediction of the N-terminal domain of CD2791 ($CD2791_{27\text{-}322}$). Prediction using PSIPRED (Jones 1999).

6.3.2 Cwp66

As mentioned in Chapter 5 specific methods (Section 6.2.1.1), the putative *C. difficile* R20291 *cwp66* gene was obtained by BLAST searching the *C. difficile* Stoke Mandeville 027 (R20291) assembly (Sanger, March 2008) with the published *C. difficile* 630 *cwp66* gene. However, the subsequent completion of the *C. difficile* R20291 genome (and publishing onto NCBI) revealed that the R20291 *cwp66* gene was in fact 25 residues longer than the Sanger Stoke Mandeville BLAST result suggested (Figure 6.3.12).

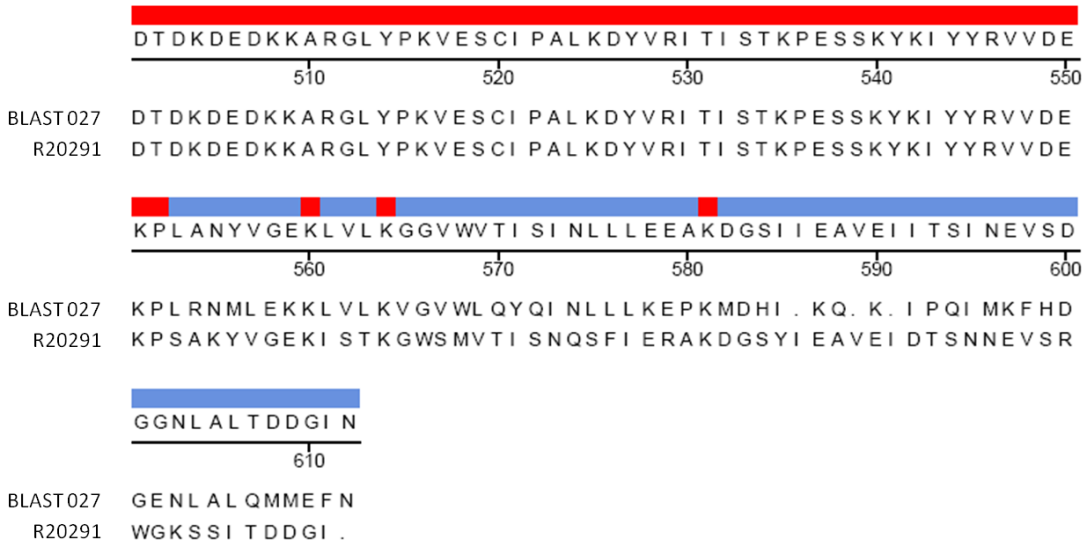


Figure 6.3.12 Alignment of transcribed putative *cwp66* from Stoke Mandeville (Sanger) BLAST (March, 2008) (labelled BLAST027) against published *C. difficile* R20291 sequence (September 2010) (Labelled R20291).

It is unknown what effect either the change or extension of residues had on the structure of Cwp66 expressed in this study. However, comparison of the C-terminal secondary structure of Cwp66 from R20291 (completed) and the putative Cwp66 R20291 sequences, reveals that the end 25 amino acids of the putative BLAST Cwp66 form a helix whereas the completed Cwp66 R20291 do not (Figure 6.3.13). The species expressed in this chapter was based on the BLAST result and as such are numbered up to I585 to represent the end of the putative gene.

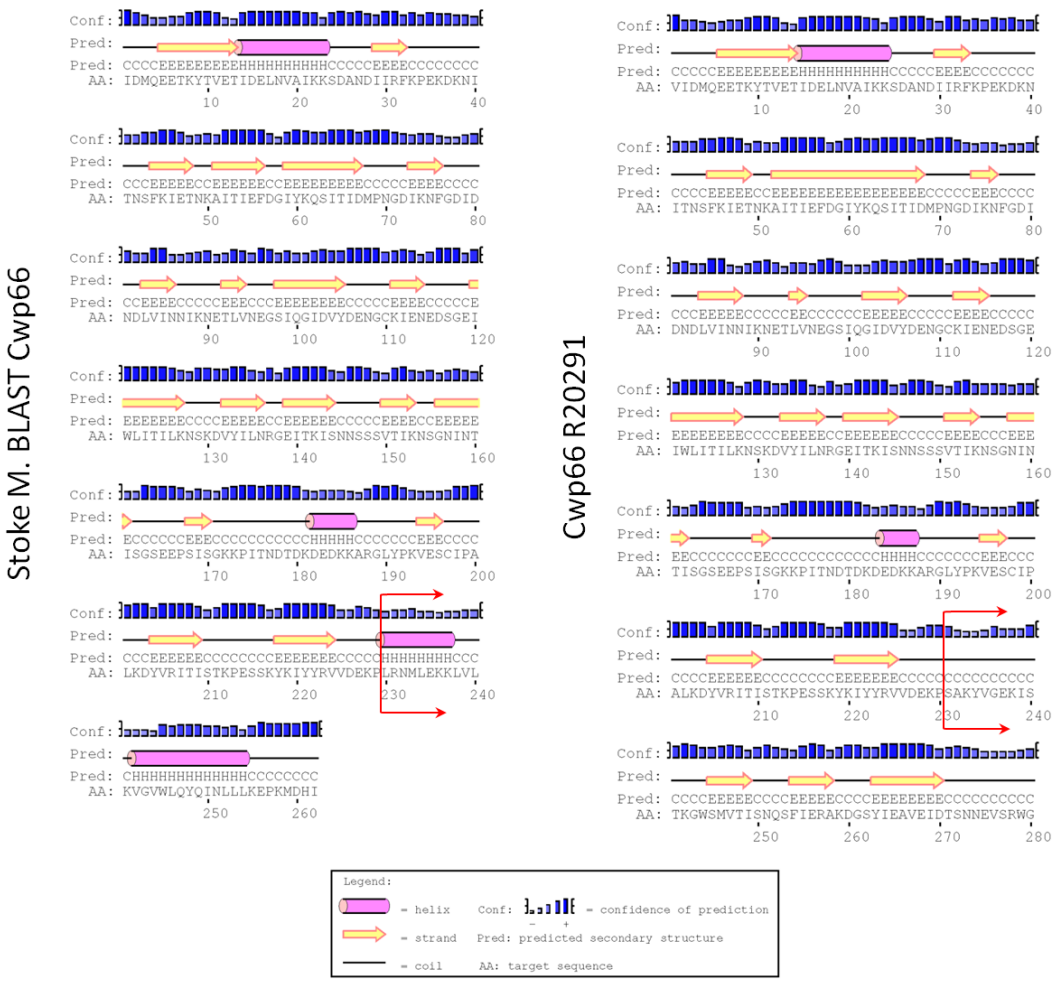


Figure 6.3.13 Secondary structure prediction of Cwp66 C-terminal regions from: putative Cwp66 (Sanger, Stoke Mandeville BLAST result) (I324-I585) and completed R20291 Cwp66 sequence (V323-I611). Arrows highlight regions different between two Cwp66 variants. Secondary structure prediction made with PSIPRED v3.0 (Jones 1999).

6.3.2.1 Expression and purification of full length rCwp66

Full length rCwp66 with His_6 and T7 tags (His.T7.rCwp66) expressed poorly and predominantly formed inclusion bodies. IMAC purification of the soluble fraction yielded a fragmented species, although purification using a Co^{2+} charged IMAC column purified a larger proportion of 66 kDa His.T7.rCwp66 than Ni^{2+} (Figure 6.3.14). The Co^{2+} IMAC eluted His.T7.rCwp66 precipitated upon dialysis into 1 M NaCl 50 mM sodium phosphate pH 7.0 for HIC purification.

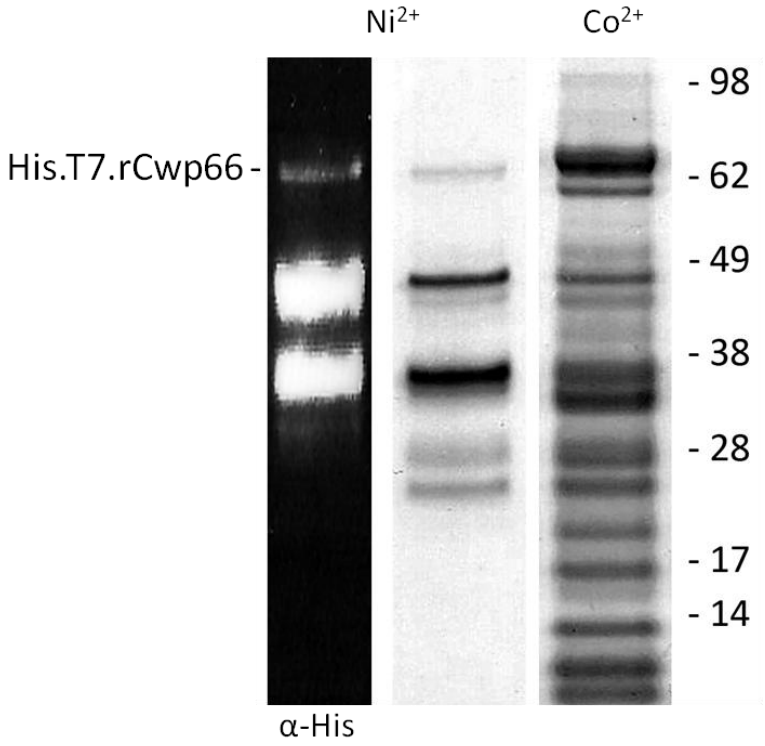


Figure 6.3.14 IMAC purification of His.T7.rCwp66. Lane 1 & 2- 5 ml HisTrap HP (nickel charged) anti-His_6 blot and IMAC elute respectively, Lane 3- 5 ml BD TALON™ (cobalt charged) media IMAC elute.

To increase solubility and ease purification of rCwp66, a GST tagged construct was assessed. Expression of this construct was largely insoluble despite optimisation of expression conditions. The limited amount of soluble GST.rCwp66 displayed progressive aggregation and precipitation during manipulation, e.g. dialysis or purification step(s), such that little to no protein could be purified.

6.3.2.2 Expression and purification of the C-terminal domain of Cwp66 ($rCwp66_{324-585}$)

The position of the N-terminal CWBDs and a propensity to form N-terminal truncates (as found in other pET vectors in this chapter) was suspected to be hindering purification. Therefore, a strategy to express only the C-terminal domain was investigated.

The 'functional' C-terminal domain of Cwp66 was expressed with only a His_6 tag (His. $rCwp66_{324-585}$). This construct expressed poorly as a 38 kDa species (~32 kDa predicted). Western blotting after IMAC revealed that the 38 kDa $His.rCwp66_{324-585}$ species only weakly bound to the IMAC column and mostly flowed through. A broad shallow elution 'peak' contained small amounts of the 38 kDa species, together with 63 and 98 kDa species, all detectable with anti-His_6 (Figure 6.3.15), suggesting $His.rCwp66_{324-585}$ may form dimers (even trimers).

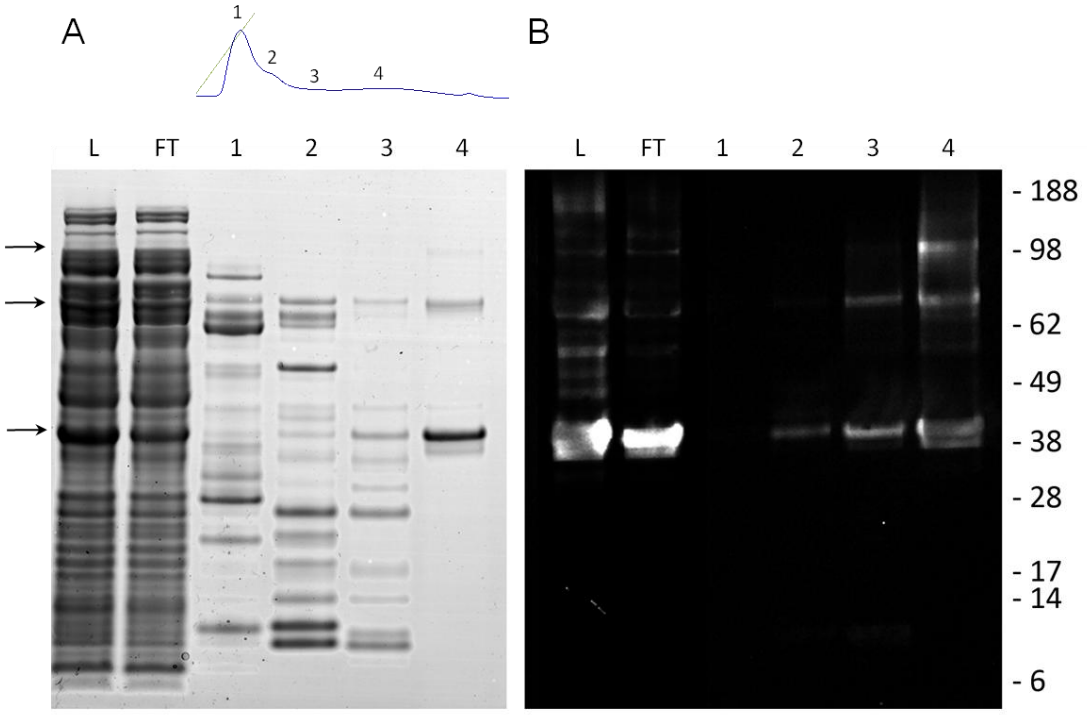


Figure 6.3.15 IMAC purification of $His.rCwp66_{324-585}$. (A) SDS-PAGE (B) Western blot with anti-His_6. L- Load *E. coli* lysate, FT- unbound flow through, E1-4 Eluted fractions.

In case the His_6 tag was occluded and to improve solubility, $rCwp66_{324-585}$ was expressed as a SUMO fusion (Figure 6.3.16). $His.SUMO.rCwp66_{324-585}$ appeared to be in the soluble fraction after cell lysis, but prior to IMAC a large amount of insoluble precipitate formed. Removal of the precipitate and application until the IMAC column became blocked or using a batch purification procedure, resulted in a small amount of $His.SUMO.rCwp66_{324-585}$ being purified. Incubation of this material

with SUMO protease cleaved the His.SUMO moiety; however, the majority of the protein was precipitated during dialysis into the cleavage buffer.

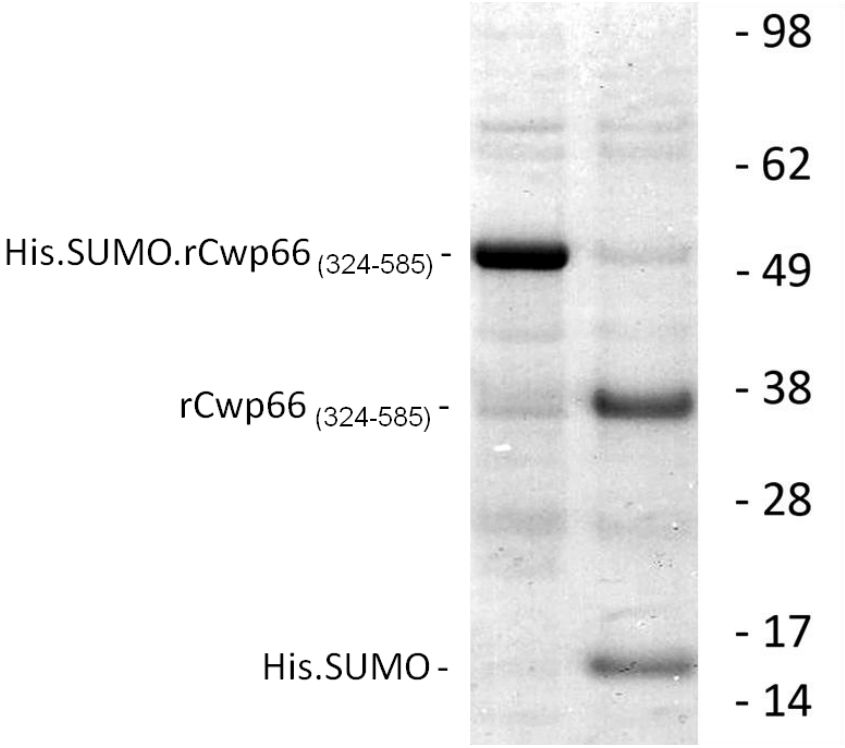


Figure 6.3.16 SUMO tagged rCwp66$_{324-585}$ purification. Lane 1- IMAC purified His.SUMO.rCwp66$_{324-585,}$ Lane 2- Incubation of 20 µg His.SUMO.rCwp66$_{324-585}$ with 10 units SUMO protease in SUMO cleavage buffer (50 mM tris, 150 mM NaCl, 0.2% igepal, 1 mM DTT pH 8.0) for 1 hr at 30 °C

The majority of the protein also precipitated upon dialysis into MonoQ IEX buffer (50 mM tris pH 8.0) after IMAC.

These results suggest that rCwp66$_{324-585}$ is a protein prone to aggregation and that this feature may be involved in its putative adhesive function *in vivo*.

6.3.2.3 Modelling of the C-terminal domain of Cwp66 from R20291 (Cwp66$_{323-611}$)

Due to the issues between the *cwp66* BLAST result from the unfinished R20291 (March 2008) and the completed R20291 *cwp66* sequence (Section 6.3.2.1),

bioinformatics' searches in this section were performed on the C-terminal 289 amino acids ($Cwp66_{323-611}$) of the finalised R20291 Cwp66 sequence, as found on NCBI.

There are currently no predicted conserved domains (CDD search (Marchler-Bauer *et al.* 2009)) or predicted protein signatures (Interpro) (Hunter *et al.* 2009) for $Cwp66_{323-611}$. BLASTP reveals limited (approx 27% identity) homology with the S-layer proteins of *Clostridium hiranonis* DSM 13275 but also finds 32% identity (K397-I475) and 26% identity (I353-L445) with cell division protein FtsQ of *Bacteroides* sp. and probable cell surface glycoprotein of Haloarchaeon respectively. Interestingly, the 96 amino acid region, similar to the cell surface glycoprotein of Haloarchaeon, corresponds to a PapD-like region, PapD is a periplasmic chaperone necessary for the assembly of pili. Performing a BLAST of the CATH database only finds matches for a 52 amino acid region with a very low E-value of 1.1 indicating an extremely limited match.

Homology modelling with SWISSMODEL (Arnold *et al.* 2006) and EsyPred3D (Lambert *et al.* 2002) confirms limited sequence similarity with known proteins, as neither could find any suitable templates. However, CPHmodels (Nielsen *et al.* 2010), using remote homology modelling, suggests T333-D609 of $Cwp66_{323-611}$ shares homology to chondroitinase B from *Pedobacter heparinus* (1OFL) (Z-score 6.9, >10 indicates a high reliability model) (Figure 6.3.17). Homology modelling therefore suggests that $Cwp66_{323-611}$ may have a parallel β sheet structure.

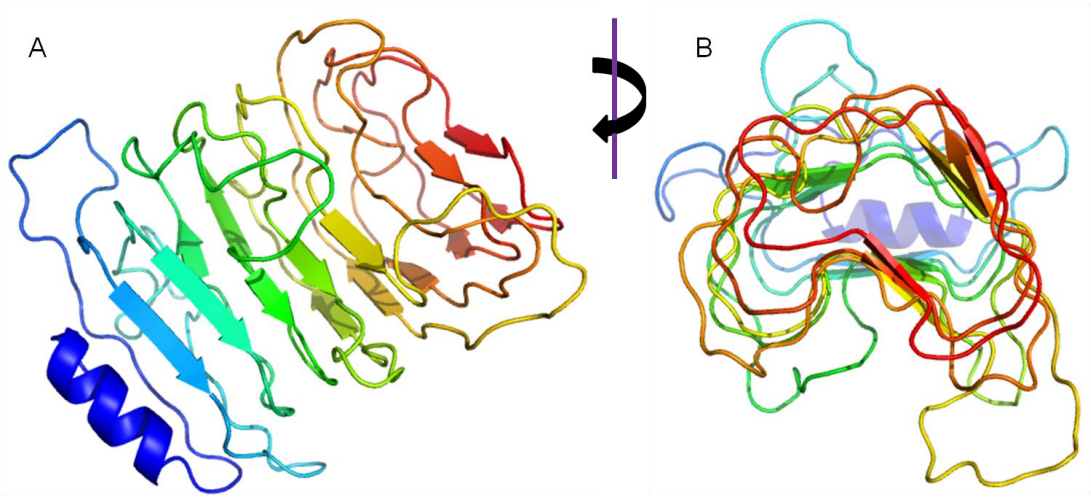


Figure 6.3.17 T333- D609 of $Cwp66_{323\text{-}611}$ modelled on chondroitinase B from *Pedobacter heparinus* (1OFL) by CPHmodels (Nielsen *et al.* 2010). (A) rotated approximately 90° to show 'hollow' core of right-handed, parallel β-helix (B). Prepared using PyMOL (www.pymol.org) (N- blue C- red).

Secondary structure prediction (PSIPRED (Jones 1999)) also suggests the C-terminal domain of Cwp66 is largely β-sheet structure (Figure 6.3.20). Given that the *cwp66* gene is variable, it is therefore likely that the variation occurs such that the structure of the protein is maintained or occurs in the loop regions. Comparison of the C-terminus of Cwp66 between 630 and R20291 reveals that the secondary structure is conserved while the amino acid sequence has 61.5% identity (Figure 6.3.20). Moreover, differing residues are found as predicted in the coil (loop) regions of the structure. This suggests that while Cwp66 is one of the few surface proteins to be genetically variable, the structure is conserved, highlighting the importance of that particular structure for Cwp66 and for role(s) that Cwp66 performs in *C. difficile*.

Threading (fold recognition) servers confirm the likely β-sheet structure; however, they suggest several methods by which the β-sheets can be arranged.

pGENThreader (Lobley *et al.* 2009) suggests a parallel β-sheet formation by alignment with the parallel β-sheet region of heme binding protease from pathogenic *E. coli* (1WXR) or alignment with the collagen adhesin (2F68) arranged

the β-sheets in sub-domains, each adopting a variant IgG-fold. Both the pGENThreader (Lobley *et al.* 2009) results are both rated high, suggesting that either arrangement is possible.

However, Phyre (Kelley & Sternberg 2009) suggests fold homology with Invasin, a bacterial integrin-binding protein (1CWV) (Hamburger *et al.* 1999) and Intimin, which together with translocated intimin receptor (Tir), mediates adhesion between mammalian cells and pathogens (1F02) (Luo *et al.* 2000). The arrangement of β-sheets in these two models is into separate domains connected by variable loop regions (Figure 6.3.18). The fact that these matches are with bacterial adhesins may suggest that if Cwp66 is an adhesin, it may adopt a fold more like these.

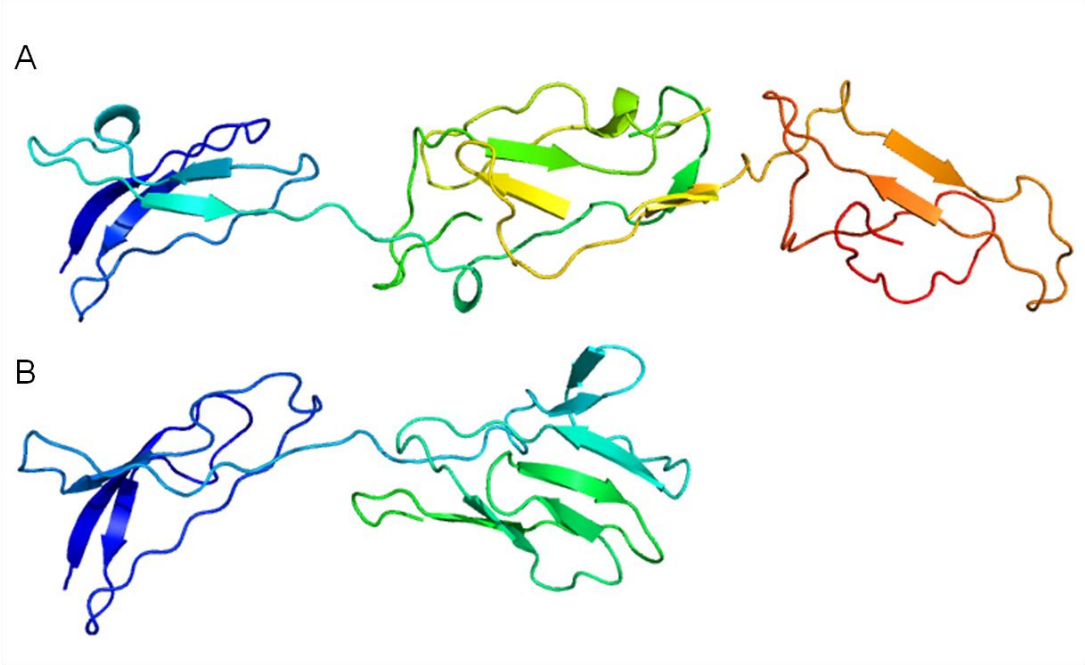


Figure 6.3.18 Phyre (Kelley & Sternberg 2009) models of $Cwp66_{323\text{-}611}$ A349-S599 on invasin (1CWV) (A) and A349-S491 on intimin (1F02) (B). Prepared using PyMOL (www.pymol.org) (N- blue C- red)(or green in (B)).

HMM-HMM homology (HHPred) (Soding *et al.* 2005) suggests that the β-sheets are arranged in a parallel β sheet structure, with homology to rhamnogalacturonase A from *Aspergillus aculeatus* (1RMG) (Petersen *et al.* 1997) or chondroitinase B from *Flavobacterium heparinum* (IDGB) (Huang *et al.* 1999) (Figure 6.3.19). Similar to

CPHmodels (Nielsen *et al.* 2010), the TM-score of HHPred for both the aforementioned alignments are approximately 0.32 and 0.29, and a TM-Score ≥ 0.4 corresponds to meaningful predictions, thus caution is applied when speculating putative structures.

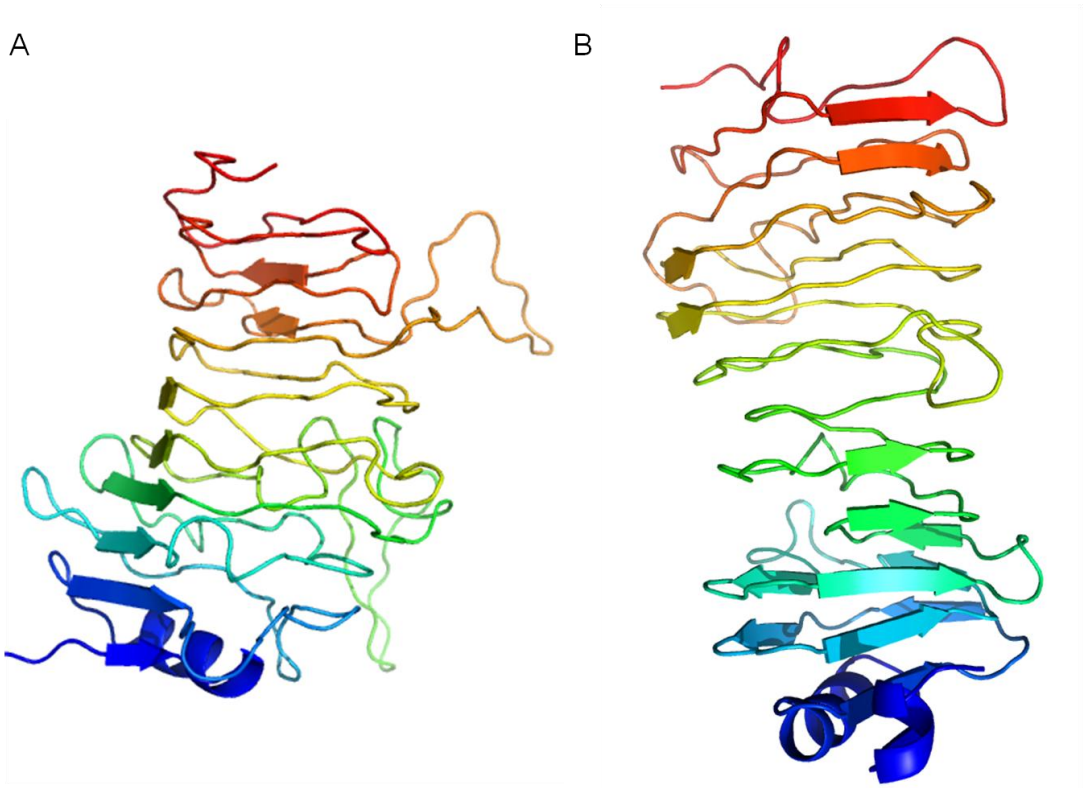


Figure 6.3.19 Cwp66$_{323\text{-}611}$ homology modelled on (A) Rhamnogalacturonase A (1RMG) (B) Chondroitinase B (1DBG). Models produced by alignment with HHPRED (Soding *et al.* 2005) then modelled in MODELLER (Eswar *et al.* 2007). Prepared using PyMOL (www.pymol.org) (N- blue C- red).

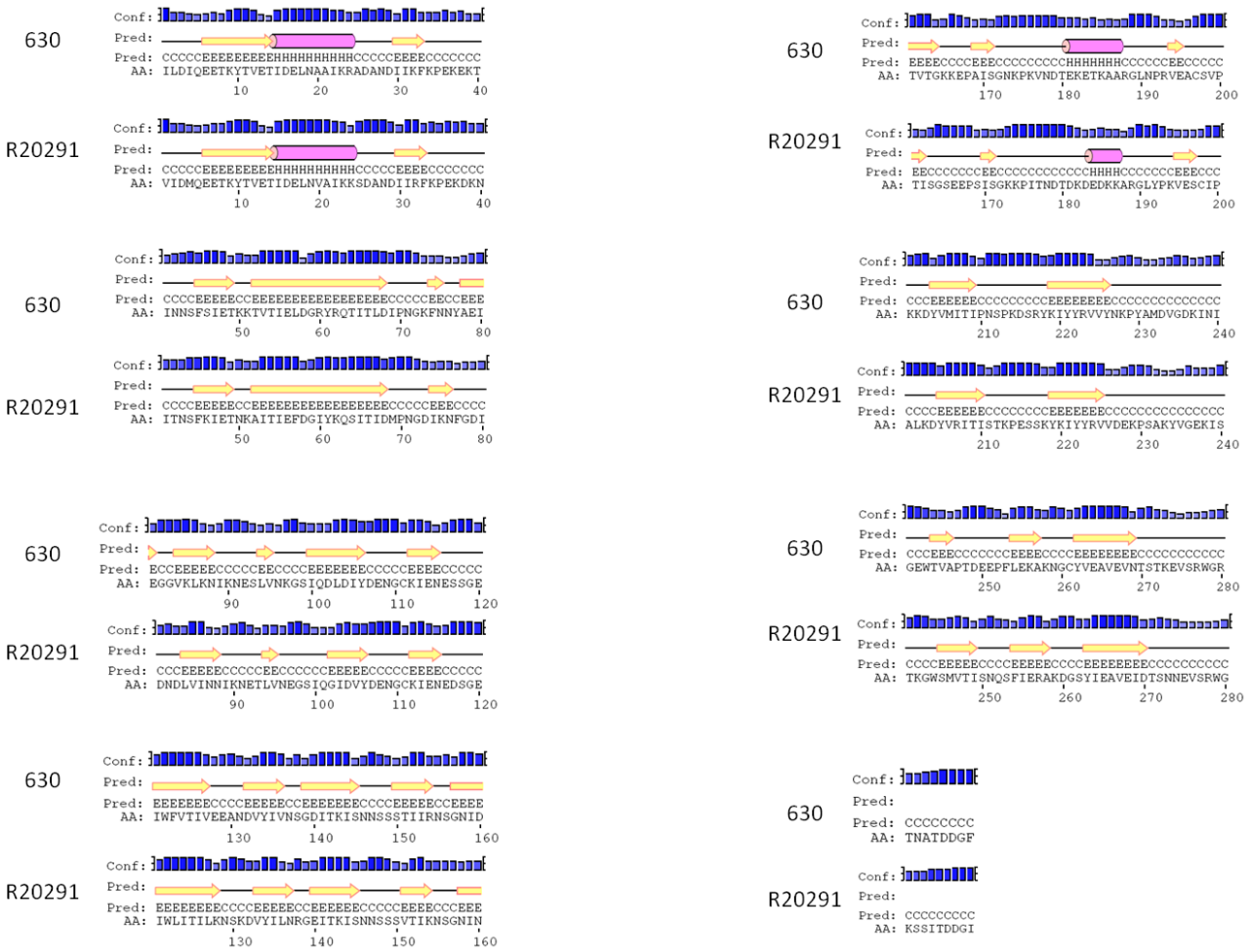


Figure 6.3.20 Comparison of secondary structure of Cwp66 C-terminal regions (residue 323+) from *C. difficile* 630 and R20291. Secondary structure predicted with PSIPRED v3.0 (Jones 1999).

6.3.3 CD2767

6.3.3.1 Expression and purification of full length rCD2767

Recombinant expression of CD2767 has not been reported to date, thus the full length protein was assessed for expression and purification. Full length His.T7.rCD2767 expressed in the soluble fraction and purified, using IMAC, as a fragmented species between 76-49 kDa (Figure 6.3.21 Lane 1), detectable with anti-His_6, suggesting N-terminal truncation to a stable 49 kDa intermediate.

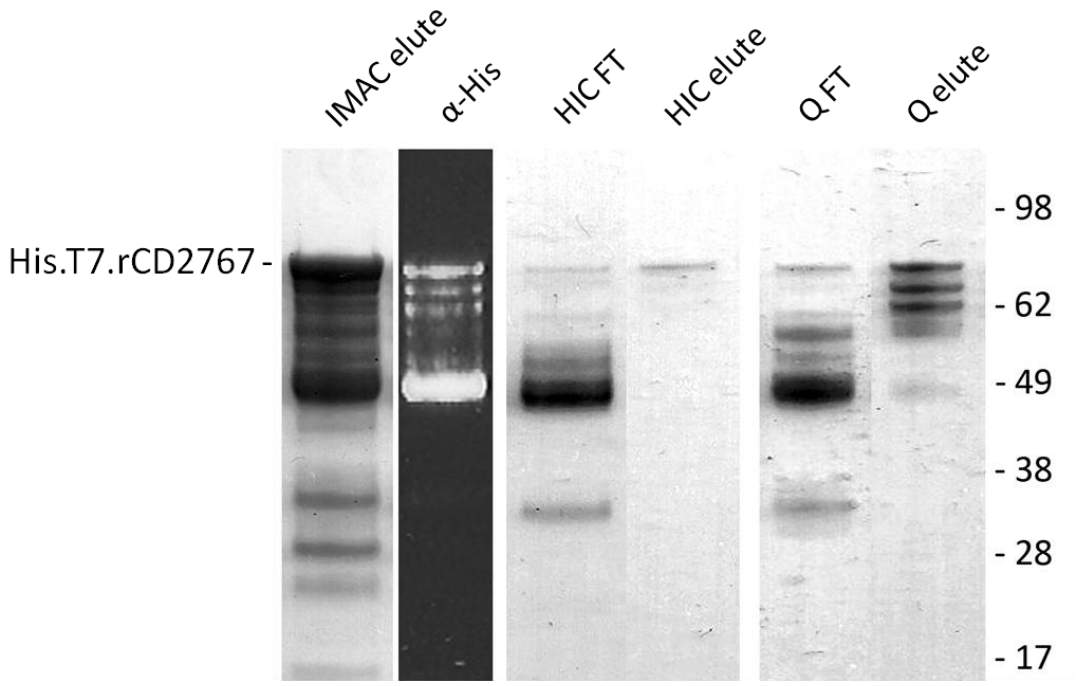


Figure 6.3.21 Purification of His.T7.rCD2767. Lane 1- Elution fraction from His.T7.rCD2767 IMAC. Lane 2- Lane1 blotted with anti-His. Lane 3 & 4- flow through and eluted material from phenyl sepharose HIC purification. Lane 5 & 6- flow through and eluted material from Q Sepharose (pH 8.0) IEX purification.

IMAC purified His.T7.rCD2767 appeared to bind irreversibly to phenyl sepharose during HIC purification with only a very small proportion being eluted (Figure 6.3.21 Lane 4), while IEX (Q, pH 8.0) purification appeared to cause triple bands (76, 66, 62 kDa) to elute (Figure 6.3.21 Lane 6), both methods appeared to somewhat remove contaminating species. A His.NusA fusion rCD2767 exhibited the same C-

terminal fragmentation and purified easily in IMAC. However, after cleavage of the His.NusA moiety with enterokinase no free rCD2767 was observed in SDS-PAGE. A GST fusion was assessed, but expressed predominantly insoluble and was not progressed further.

6.3.3.2 Expression and purification of rCD2767$_{27\text{-}401}$

Due to the extensive N-terminal truncation/degradation and problems obtaining pure protein with full length rCD2767, the expression and purification of the N-terminal 'functional' domain (corresponding to residues 27-401 of the full length CD2767) containing the predicted glycosidase catalytic core, was assessed.

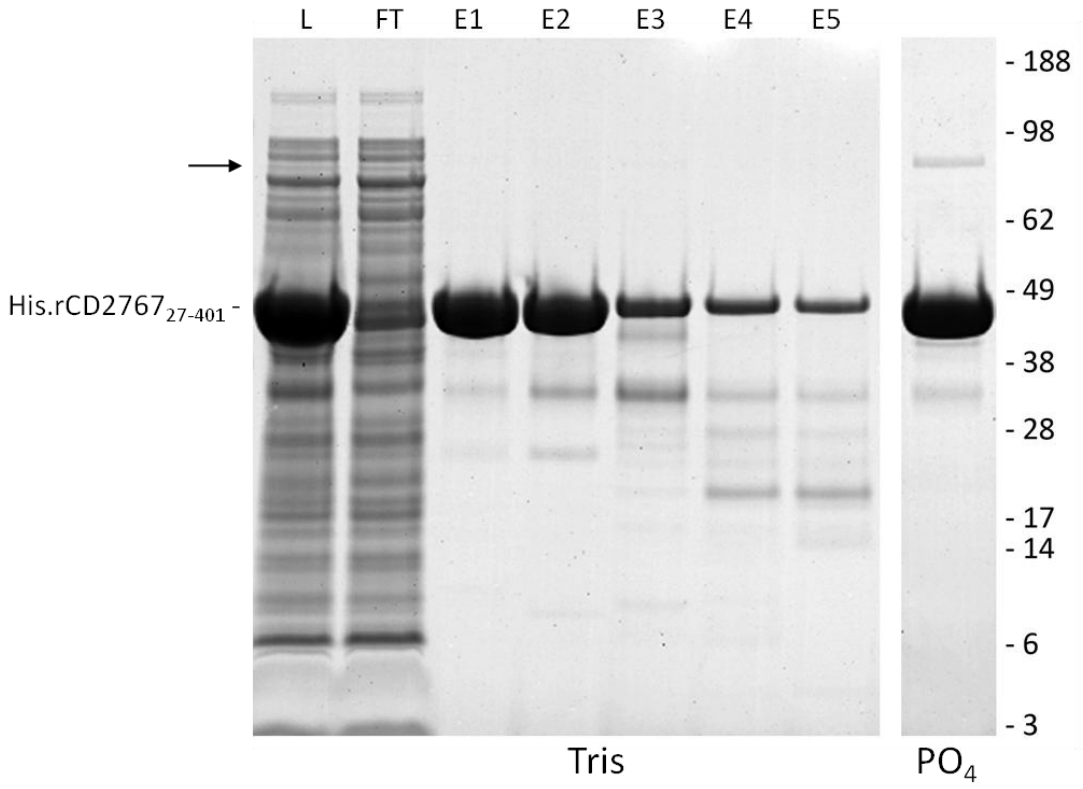


Figure 6.3.22 IMAC purification of His tagged rCD2767$_{27\text{-}401}$. Lanes 1-7 - tris based IMAC buffers (L – *E. coli* lysate, FT- unbound material, E1-5 eluted fractions from early (E1) and late (E5) in eluted peak) Lane 8 - sodium phosphate (monobasic) based IMAC buffers. Arrow denotes approximate dimer MW.

Using only an N-terminal His_6 tag (His.rCD2767$_{27-401}$), IMAC purification yielded a highly pure 47 kDa species in SDS-PAGE in one step, particularly early in the elution peak (Figure 6.3.22). Only this species was detected by western blot with anti-His_6, confirming both its identity and that other eluted bands were not N-terminal truncates.

SEC of this material gave a single peak corresponding to a monomeric protein of approximately 28 kDa (~42 kDa predicted). When purifying His.rCD2767$_{27-401}$ with tris based IMAC buffers, a single species was predominant with some lower MW contaminants (as seen in Figure 6.3.22). However, IMAC purification using phosphate based IMAC buffers; an 87 kDa species was present, which blotted with anti-His_6 suggesting dimerisation of His.rCD2767$_{27-401}$.

Fractions from early in the tris based IMAC elution peak, >90% SDS-PAGE 'pure' His.rCD2767$_{27-401}$ (as measured by densitometry) was pooled, concentrated then dialysed into 50 mM tris 150 mM NaCl pH 8.0 prior to crystallisation. His.rCD2767$_{27-401}$ could be concentrated (by centrifugal spin column) to 120.2 mg/ml (by A_{280}) or 167 mg/ml (by Bradford Assay using BSA as the standard) (both quantifications used a 1:500 dilution), suggesting a very soluble protein. During crystallisation set-up the nanodispensing robot ball-bearing valve system remained open, suggesting an increased viscosity corresponding to high protein concentrations. Fusion of rCD2767$_{27-401}$ to a C-terminal toxin B fragment appeared to increase the toxin fragments solubility to the same extent as commercial thioredoxin and NusA fusion tags, confirming rCD2767$_{27-401}$ had soluble properties.

Mass spectrometric analysis suggested, that despite the loss of the N-terminal methionine, that the observed molecular weight (44644 Da + 131 Da (Met) = 44775 Da) matched the predicted molecular weight (44775 kDa). Storage of the putatively highly concentrated His.rCD2767$_{27-401}$ for 3 months at 4 °C resulted in a decrease in quantified protein concentration. Both the A_{280} and Bradford assay concentrations were approximately 50% figures obtained from freshly purified His.rCD2767$_{27-401}$. Analysis of the stored material by SDS-PAGE did not reveal

additional bands, suggesting precipitation was lowering the overall protein quantity rather than degradation.

6.3.3.3 Identification of functional enzymatic activity for rCD2767$_{27\text{-}401}$

As CD2767 has not been characterised to date, in order to determine if His.rCD2767$_{27\text{-}401}$ was correctly folded, an activity assay was required. To develop an assay, a substrate/ligand for the putatively glycosidase catalytic core of CD2767$_{27\text{-}401}$ needed to be identified.

A protein signature domain search with InterProScan (Hunter *et al.* 2009) suggests CD2767$_{27\text{-}401}$ is a TIM beta/alpha barrel O-glycoside hydrolase. An O-glycoside is a sugar linked to a non-carbohydrate via an O-glycosidic bond (in reference to the glycosidic oxygen that links the glycoside to the aglycone or reducing end sugar). Using a NCBI Conserved domain search (Marchler-Bauer *et al.* 2009), CD2767$_{27\text{-}401}$ had homology to alpha amylase (Family 13), melibase or the glycosyl hydrolase Family 31 (GH31) including alpha-glucosidase (glucoamylase and sucrase-isomaltase), alpha-xylosidase, 6-alpha-glucosyltransferase, 3-alpha-isomaltosyltransferase and alpha-1,4-glucan lyase.

BLASTP revealed that CD2767$_{27\text{-}401}$ shares high identity (≥96%) with surface proteins of *Clostridium hiranonis* DSM 13275 (EEA85382.1), *Peptostreptococcus stomatis* DSM 17678 (EFM64818.1) and *P. anaerobius* 653-L (EFD05798.1), and shares 91% identity with the FenI protein of *Streptosporangium roseum* DSM 43021 (ACZ88859.1), suggesting the role CD2767 may performs is conserved across several bacterial species. However, BLASTP using the Protein Data Bank (PDB), only finds a small region of CD2767$_{27\text{-}401}$ (L37 – D169) that shares 25% identity with a glycosyltrehalose trehalohydrolase from *Sulfolobus Solfataricus* (alpha-amylase (Family 13)) (1EHA) (Feese *et al.* 2000; Yoneda *et al.* 2003) or a 83 aa region (Y137-V220) which shares 27% identity with a glycogen branching enzyme 1,4-alpha-D-glucan 6-glucosyl-transferase from *Mycobacterium tuberculosis* H37rv (3K1D).

As a result of these findings, the following assays were assessed with rCD2767$_{27\text{-}401}$ to identify any enzymatic activity/substrate for CD2767:

- Alpha amylase activity (starch hydrolysis – Phaedebas© test)
    - From conserved domain search
- Trehalase assay (based on detection of liberated glucose)
    - Based on the limited amino acid sequence identity to glycosyltrehalose trehalohydrolase from *Sulfolobus Solfataricus*
- O-glycosidase assay (hydrolysis of p-nitrophenyl galacto-N-bioside)
    - from InterProScan results
- β- glucosidase assay (hydrolysis of p-nitrophenyl-β-D-glucopyranoside)
    - A common TIM barrel glycosidase

Where necessary rCD2767$_{27\text{-}401}$ was re-purified using buffers, or in the absence of buffers may have affected results, e.g. chloride ions activate, while tris ions inhibit alpha amylase activity (Aghajari *et al.* 1998) yet chloride ions inhibit O-glycosidase activity (Umemoto *et al.* 1977).

No activity was observed in the assays tested. These data suggest that the substrate of CD2767 is not one of those assessed in this chapter, as glycosidases can be specific for their substrate, or that rCD276727$_{\text{-}401}$ was not correctly folded or requires a metal co-factor for activity. The lack of activity observed by rCD2767$_{27\text{-}401}$ may reveal that, putatively like recombinant Cwp84, the CWBDs are required for spatial positioning for enzymatic activity.

6.3.3.4 Modelling CD2767 domain

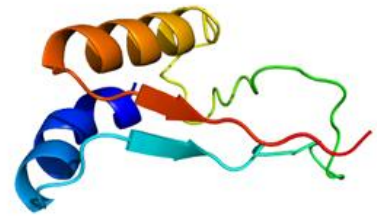
Protein signature searches suggested that CD2767$_{27\text{-}401}$ had sequence homology with TIM barrel structures which have β-sheets forming a central core around which α-helices are arranged (Figure 6.1.10). To further identify any putative substrate, the putative glycosidase domain of CD2767 (CD2767$_{27\text{-}401}$) was entered into a number of tertiary structure prediction servers.

Using homology modelling (the assumption that two homologous amino acid sequences will share very similar structures) appears to provide little more information about $CD2767_{27\text{-}401}$ than BLASTP results (Section 6.3.3.3). Performing a BLAST of the CATH database identifies the 133 aa region of $CD2767_{27\text{-}401}$ is similar to the glycosyltrehalose trehalohydrolase (1EH9 and 1EHA) (whose structure was solved by multiple isomorphous replacement) (Feese *et al.* 2000).

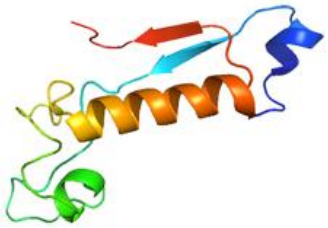
SWISSMODEL (Arnold *et al.* 2006) found six templates upon which $CD2767_{27\text{-}401}$ could be aligned and while all had low sequence identity, with decreasing identity the models produced were increasingly TIM barrel-like (Figure 6.3.24).

Geno3D (Combet *et al.* 2002) modelled the 132 aa region of limited identity between $CD2767_{27\text{-}401}$ (L37 – D169) and the glycosyltrehalose trehalohydrolase from *Sulfolobus Solfataricus* (1EHA or 1EH9). The model produced appears to correspond to some of the α-helices surrounding the barrel core (Figure 6.3.23).

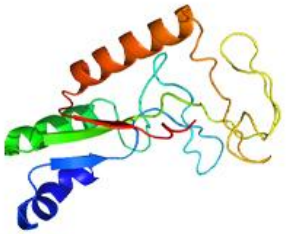
Figure 6.3.23 Geno3D (Combet *et al.* 2002) homology model of $CD2767_{27\text{-}401}$ (L37 – D169) (Blue) aligned to glycosyltrehalose trehalohydrolase from *Sulfolobus Solfataricus* (1EHA) (Feese *et al.* 2000)(Grey). Models superposed in Coot (Emsley *et al.* 2010) and prepared using PyMOL (www.pymol.org).



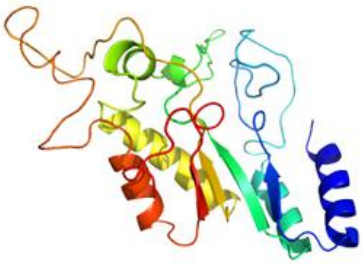
modelled residue range:	31 to 104
based on template	1gjwA (2.10 A)
Sequence Identity [%]:	18.391
Evalue:	6E-09



modelled residue range:	84 to 172
based on template	2fhfA (1.65 A)
Sequence Identity [%]:	18
Evalue:	2.3E-10

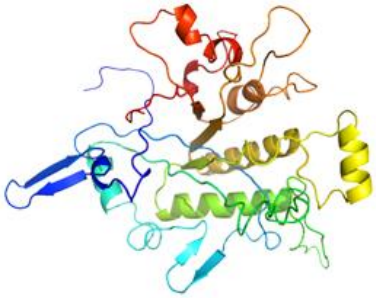


modelled residue range:	31 to 173
based on template	2aaaA (2.10 A)
Sequence Identity [%]:	17.647
Evalue:	4.2E-11



modelled residue range:	28 to 236
based on template	3mi6D (2.70 A)
Sequence Identity [%]:	16.818
Evalue:	3.2E-17

modelled residue range:	37 to 346
based on template	2gsjA (1.73 A)
Sequence Identity [%]:	12.693
Evalue:	6.4E-05



modelled residue range:	3 to 308
based on template	3bxwB (2.70 A)
Sequence Identity [%]:	7.837
Evalue:	2E-07

Figure 6.3.24 SWISS-MODEL (Arnold *et al.* 2006) alignments of $CD2767_{27\text{-}401}$ and corresponding models of those regions. The lower the E-value, or the closer it is to zero, the more "significant" the match is.

Similar to the BLAST result (Section 6.3.3.3), ESyPred3D (Lambert *et al.* 2002) finds $CD2767_{27\text{-}401}$ S38-V318 has limited identity with the 1,4-alpha-glucan-branching enzyme from *Mycobacterium tuberculosis* H37RV (3K1D) sufficiently high to produce a model. The region aligned by ESyPred3D (Lambert *et al.* 2002) corresponds to an outer region of the protein away from the barrel core (Figure 6.3.25).

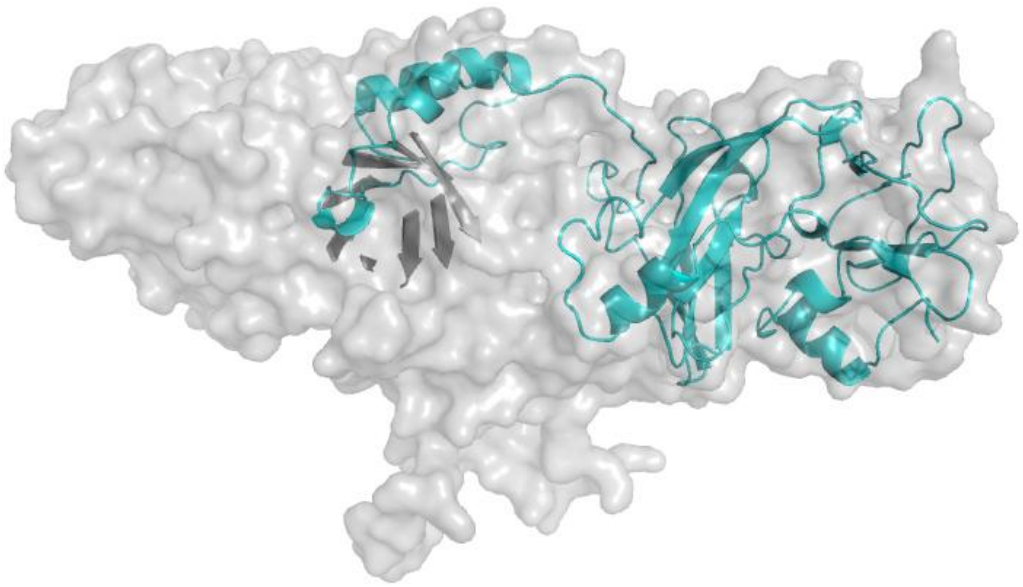


Figure 6.3.25 ESyPred3D (Lambert *et al.* 2002) tertiary model of $CD2767_{27\text{-}401}$ on 3K1D. $CD2767_{27\text{-}401}$ (blue), 3K1D surface (white) with TIM barrel highlighted (grey). Prepared using PyMOL (www.pymol.org).

The TIM barrel nature of $CD2767_{27\text{-}401}$ appears to be confirmed upon tertiary structure prediction by threading methods (fold recognition). This method is particularly applicable for glycosidases, as the TIM barrel fold is often more conserved than the sequence.

The Phyre server (Kelley & Sternberg 2009) suggests $CD2767_{27\text{-}401}$ folds belong to the amylase catalytic domain family. Proteins with similar folds are *E. coli* 1,4-alpha-glucan branching enzyme (1M7X) (Abad *et al.* 2002) (E-value 3.1e-17), isomaltulose synthase (PalI) of Klebsiella sp. LX3 (1M53) (Zhang *et al.* 2003) (E-value 4.9e-17) and *B. cereus* oligo-1,6-glucosidase (1UOK)(Watanabe *et al.* 1997) (E-value 7.4e-17).

HHPred (Soding *et al.* 2005) also finds a 213 aa region has folds common to the glycosyl transferase *E. coli* 1,4-alpha-glucan branching enzyme (1M7X), alpha-galactosidase (3MI6) and cyclomaltodextrinase (1EA9). Models produced using the HHPred alignments with MODELLER (Eswar *et al.* 2007) suggest a conserved Aspartic acid (D195) at the bottom of the core at the end of a β-sheet in each of the models of the above enzymes (Figure 6.3.26 marked in pink). Homology models are only models and are biased on the template used and cannot define the nature of arrangement or conformation of the protein molecules, as defined by experimental methods.

pGENThreader (Lobley *et al.* 2009), suggests matches ranked 'certain' with endo-beta-1,4-galactanases (involved in pectin degradation) from *Bacillus licheniformis* (1UR4), the fungi *Humicola insolens* (1HJQ) and A*spergillus aculeatus* (1FOB), as expected, all are TIM barrels and two of which (1UR4 and 1FOB) require calcium for activity.

In summary, the N-terminal domain of CD2767 shows no significant sequence similarity to any known glycoside hydrolase. Fold prediction aligns CD2767 with other glycosidases, all with a TIM barrel, suggesting that this particular fold family be assessed during data processing or model building. However, given the substrate specificity (despite structural similarities as pointed out by Lee *et al.* (2002a)) specific ligand determination, as attempted in section 6.3.3.3, was unsuccessful.

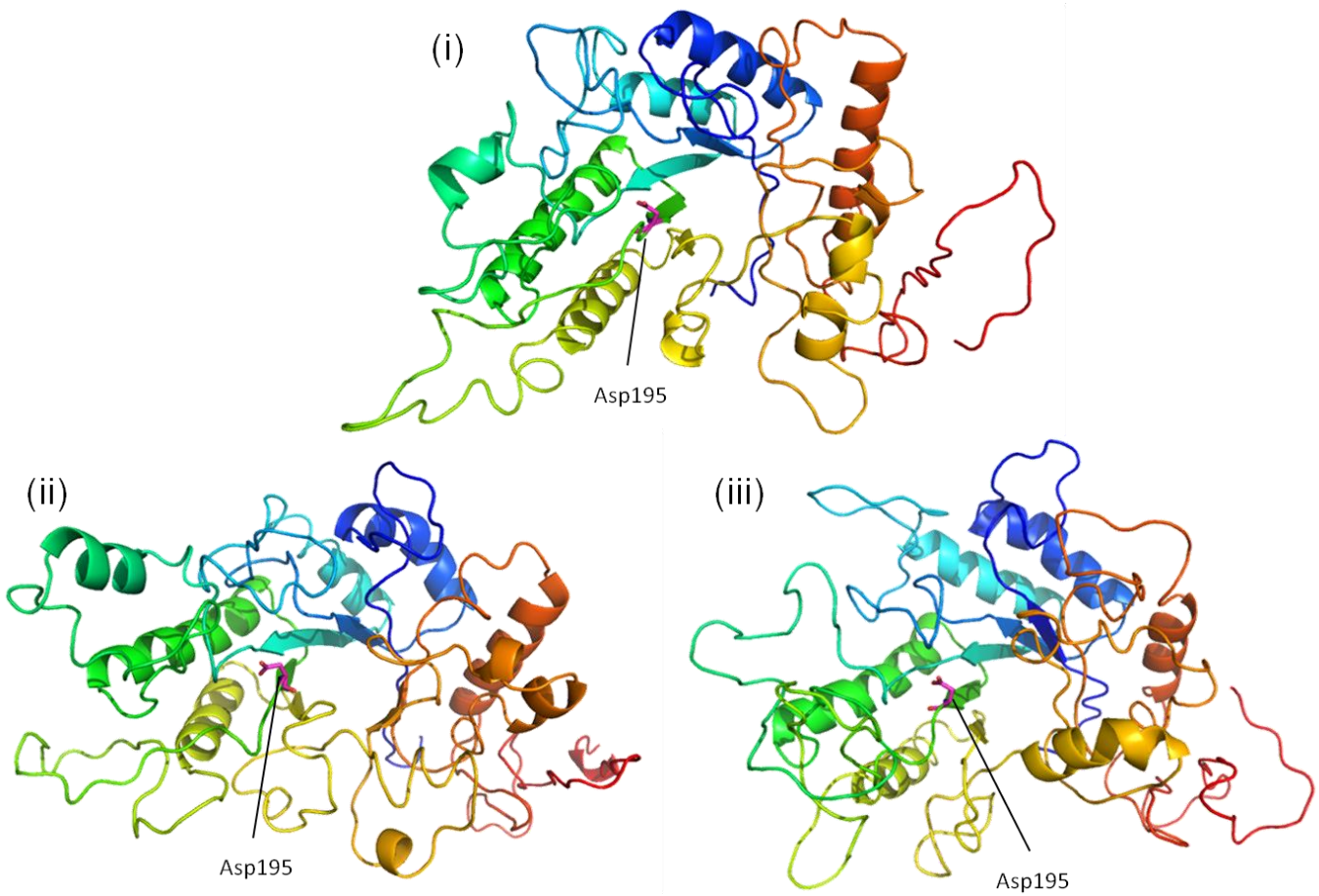


Figure 6.3.26 Predicted structures of $CD2767_{27\text{-}401.}$ Homology modelled on (i) alpha-galactosidase (melibase) from *Lactobacillus brevis* (3MI6) (ii) Cyclomaltodextrinase (1EA9) (iii) *E. coli* glycogen branching enzyme (1M7X). Conserved aspartic acid shown in pink. Models produced by MODELLER (Eswar *et al.* 2007) after alignment with HHPred (Soding *et al.* 2005). Prepared using PyMOL (www.pymol.org) (N- blue C- red).

6.3.3.5 Crystallisation of rCD2767$_{27\text{-}401}$

6.3.3.5.1 Crystallisation set-up

Sitting drop vapour diffusion crystallisation screens were setup with the highly concentrated His.rCD2767$_{27\text{-}401}$, after 3 months no observed hits could be seen, nor could any precipitation in any condition. However, after 4 months there were visible hits in several conditions (Table 6.3.4). Most notably a large crystal in one condition in the Structure Screen 1 & 2 screen (Figure 6.3.27). A preparation of His.rCD2767$_{27\text{-}401}$ concentrated to 6 mg/ml (A_{280}) and dialysed into 50 mM MES 150 mM NaCl pH 6.0 had small crystals in a well at 2:1 protein:reservoir concentration in the same D10 screen condition after a similar timescale. This suggests that 0.05 M potassium dihydrogen phosphate, 20 % w/v PEG 8000 is the likely condition by which one can crystallise His.rCD2767$_{27\text{-}401}$.

Figure 6.3.27 His.rCD2767$_{27\text{-}401}$ crystal. Crystal formed in sitting drop vapour diffusion crystallisation of His.CD2767$_{27\text{-}401}$ in 0.05 M potassium dihydrogen phosphate, 20 % w/v PEG 8000 with 1:1 protein: reservoir ratio.

This crystal was delicate when probed; however, a portion could be picked up with a loop and loaded onto the goniometer, this portion diffracted to approximately 2 Å using the Diamond synchrotron X-ray source (Figure 6.3.28).

Table 6.3.4 Sitting drop crystallisation conditions which demonstrated varying degrees of crystallisation His.rCD2767$_{27-401}$.

Reservoir conditions						
Screen	**Well**	***Buffer/Salt***	***Buffer/Salt***	***pH***	***Precipitant***	**Result**
SS 1 & 2	D10	0.05 M potassium dihydrogen phosphate	None	None	20 % w/v PEG 8000	Figure 6.3.27
PACT premier	A5-6	0.1 M SPG buffer	None	8.0-9.0	25 % w/v PEG 1500	Tiny Sea Urchins
	B5-6	0.1 M MIB buffer	None	8.0-9.0	25 % w/v PEG 1500	As above
	C6	0.1 M PCTP buffer	None	9.0	25 % w/v PEG 1500	As above
	D6	0.1 M MMT buffer	None	9.0	25 % w/v PEG 1500	As above
	C10	0.2 M magnesium chloride	0.1 M HEPES	7.0	20 % w/v PEG 6000	As above
	D10	0.2 M magnesium chloride	0.1 M tris	8.0	20 % w/v PEG 6000	As above
Screen	**Well**	***Ligand stock***	***Buffer stock***	***pH***	***Precipitant stock***	**Result**
Morpheus	C9	0.09M nitrate phosphate sulphate (NPS)	0.1M Buffer 3	8.5	30.0% P550MME_P20K	Tiny Sea Urchins

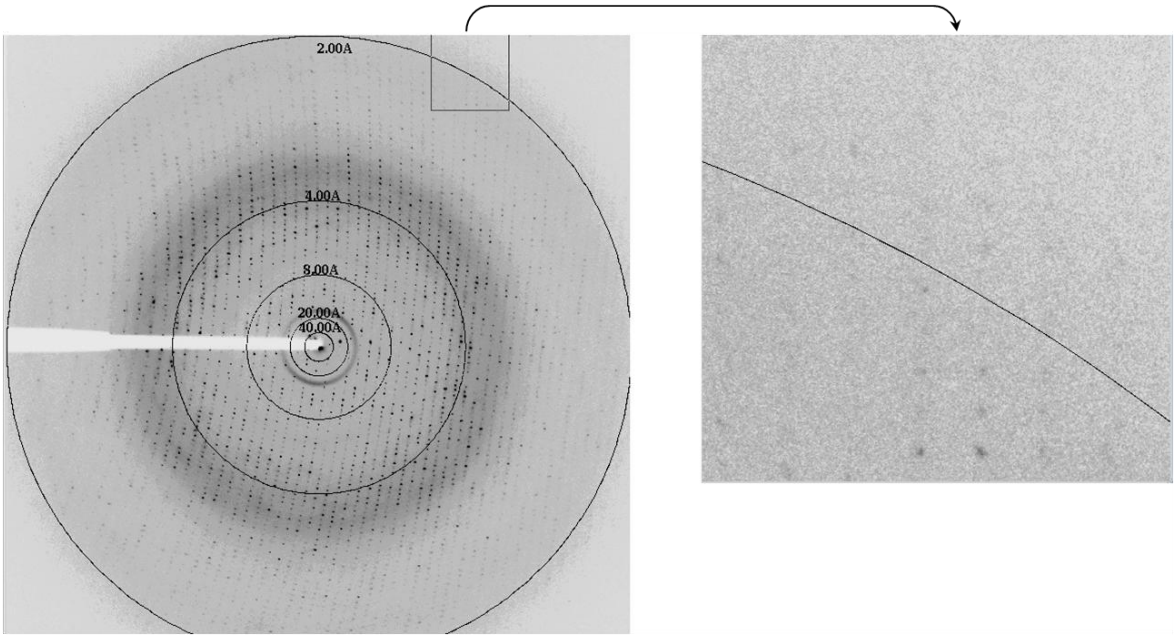


Figure 6.3.28 X-ray diffraction pattern of His.CD2767$_{27\text{-}401}$ crystal from Figure 6.3.27.

6.3.3.5.2 Crystal data processing

X-ray diffraction patterns of the His.CD2767$_{27\text{-}401}$ crystal were initially processed with iMOSFLM (Leslie 2006) and scaled using SCALA (CCP4 1994) (as per section 6.2.5.4). Initial indexing resulted in a space group ambiguity between: C-centered orthorhombic (C222 & C222$_1$), primitive monoclinic (P2 & P2$_1$), C-centered monoclinic (C2) and primitive triclinic (P1) (Table 6.3.5).

Matthew's coefficient (Matthews 1968) was calculated to estimate both the number of molecules per asymmetric unit and resulting % volume occupied by solvent. This is derived from the unit cell dimensions (and angles), the molecular weight of CD2767$_{27\text{-}401}$ and number asymmetric of units in the unit cell (i.e. the number of symmetry operators in the space group). Due to the high solvent content (which is typically between 27-65% (Matthews 1968)), and the high R_{merge} value, C-centered orthorhombic (C222 & C222$_1$) space groups were unlikely. While, R_{merge} was lower in P2$_1$ (P2) and P1 space groups, 6 molecules per asymmetric unit were estimated for the P1 space group whereas a more likely 3 molecules per asymmetric unit for P2 and P2$_1$, making the P1 space group unlikely. POINTLESS in iMOSFLM (Leslie 2006), which scores all the possible Laue groups consistent with the crystal class by matching potential symmetry equivalent reflections,

also suggested the $P2_1$ space group. Preliminary molecular replacement trials were attempted using primitive monoclinic $P2_1$ space group.

6.3.3.5.2.1 Molecular replacement (MR) trials

Due to the problems outlined in Section 6.3.3.3 and 6.3.3.4, i.e. a lack of proteins with sufficient sequence identity to $CD2767_{27\text{-}401}$, finding a model to use in molecular replacement was difficult. Various models from homology modelling servers, e.g. Phyre (Kelley & Sternberg 2009) and SWISS-MODEL (Arnold *et al.* 2006), were tried with molecular replacement program PHENIX (Adams *et al.* 2010) to assess the feasibly of solving the structure. All models were taken through an initial round of refinement, despite low translation function Z-score (TF Z-scores) (typically <6 in the "unlikely" category (TF-Z >8 = definitely solved)), yielding phase error (lower numbers suggest accurate structures) and figure of merit (FOM) statistics (higher numbers suggest accurate structures) (Table 6.3.6).

A model from Phyre (Kelley & Sternberg 2009) and i-Tasser (Zhang 2008) was assessed, but both phase error and the FoM were too high and low respectively, probably due to complexity of the tertiary structure prediction models, e.g. incorrect loops. Therefore, models produced by homology modelling server SWISSMODEL (Arnold *et al.* 2006), which contained only regions of higher homology (identity or similarity), were progressed further.

Initially, searching for 1 molecule in the asymmetric unit, model 5 of SWISSMODEL (Arnold *et al.* 2006) (based on a portion of 2GSJ See Figure 6.3.24) gave the lowest phase error and highest figure of merit (0.39). A polyalanine model (to remove side chain interference) did not give satisfactory statistics. Molecular replacement was assessed using SWISSMODEL model 5 but with 2 molecules per asymmetric unit, but was not successful (PHENIX (Adams *et al.* 2010) gave an error of "asymmetric unit too full" with the suggested solution being to search for 1 molecule).

In summary, due to the low sequence similarity of the $CD2767_{27\text{-}401}$ with known proteins (in the protein databand), it is unlikely that the structure will be solved by molecular

replacement. Given that the glycosidases often have conserved structure but not sequence, this is not surprising. Further work crystallising the CD2767 N-terminal domain with heavy metal derivatives should result in the solution of the structure.

6.3.4 Summary

Selected *C. difficile* surface proteins that represented characterised (Cwp66), partially characterised (CD2791) and uncharacterised (CD2767) proteins were trialled for their ability to be expressed and purified for X-ray crystallographic studies. It was found that the CWBD hindered purification and upon removal, both CD2791 and CD2767 were purified to sufficient purity. The N-terminal domain of CD2767 appeared to be very soluble and the identification of crystallisation conditions and an initial data set to 2.0 Å, means its structure should be solvable (although not by molecular replacement).

Table 6.3.5 Potential space groups for the His.CD2767$_{27-401}$ crystal data.

	$C222_1$		C222		C2		C2(penalty 3)		P2		$P2_1$		P1 (penalty 1)	
	Overall	Outershell	Overall	Outershell	Overall	Outershell	Overall	Outershell	Overall	Outershell	Overall	Outershell	Overall	Outershell
Low Resolution Limit (Å)	54.16	2.11	54.15	2.11	54.06	2.11	54.07	2.11	54.06	2.11	54.06	2.11	54.13	2.11
High Resolution limit (Å)	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
R_{merge}	0.513	0.720	0.502	0.656	0.489	0.666	0.468	0.609	0.135	0.538	0.135	0.538	0.100	0.439
Completeness	99.1	98.1	98.3	97.8	97.9	95.9	94.6	93.1	91.4	82.9	91.4	82.9	84.3	75.6
a (Å) / α (°)	122.41	90	122.39	90	122.32	90	180.84	90	109.13	90	109.13	90	61.23	111.84
b (Å) / β (°)	181.18	90	181.16	90	180.8	89.98	122.34	90.08	61.2	111.85	61.2	111.85	109.23	90.08
c(Å) / γ (°)	61.26	90	61.25	90	61.2	90	61.22	90	109.15	90	109.15	90	109.29	89.93
Volume of Unit cell ($Å^3$)		1358639.2		1358045.6		1353465.9		1354429.2		728986.2		728986.2		730948.2
Divided by number of asymmetric units in the unit cell (Z)		169829.9 (8)		169755.7 (8)		338366.5 (4)		338607.3 (4)		364493.1 (2)		364493.1 (2)		730948.2 (1)
Divided by MW (44,775 Da)		3.79		3.79		7.56		7.56		8.14		8.14		16.32
Molecules/asymmetric unit*		1		1		1		3		3		3		6
V_m		3.79		3.79		7.56		2.52		2.71		2.71		2.72
V'_p (≈1.23/Vm)		0.32		0.32		0.16		0.49		0.45		0.45		0.45
V_{solv} (1-V'_p)		68%		68%		84%		51%		55%		55%		55%

Penalty figures in brackets refer to a spacegroup solution but with a higher penalty.

a, b, c refer to cell dimensions whereas α, β, γ refer to cell angles.

R_{merge} - A statistical measure of how well the X-ray diffraction data scale together (good-quality structures have low values).

* Most likely number of molecules/asymmetric unit.

Vm - Matthews coefficient (typically between 1.68 - 3.53).

V'p – Fraction of the total crystal volume occupied by protein.

V_{solv} - Putative percentage solvent in the unit cell.

Table 6.3.6 Molecular replacement statistics from PHENIX (Adams *et al.* 2010) using various tertiary CD2767 structure prediction models with $P2_1$ space group.

	SWISSMODEL Model 5		SWISSMODEL Model 6		Phyre d1m7xa3 model		PolyAla Model 5 2gsj		2gsj		I-TASSER (Model 2)	
	initial	*final*	*initial*	*final*	*initial*	*final*	*initial*	*final*	*initial*	*final*	*initial*	*final*
R_{work}	0.817	0.495	0.828	0.493	0.577	0.490	0.575	0.515	0.572	0.486	0.577	0.483
R_{free}	0.763	0.526	0.768	0.522	0.574	0.525	0.571	0.537	0.551	0.521	0.582	0.540
Figure of Merit (FOM)		0.390		0.210		0.140		0.250		0.330		0.250
Phase Error		61.640		74.960		80.550		72.400		66.750		73.020

Definitions:

R_{work}- a measure of the agreement between the crystallographic model and the experimental X-ray diffraction data (typically between 15% to 25% for accurate structures).

R_{free} (Brunger 1992)- measures the agreement between observed and computed structure factor amplitudes for a 'test' set of reflections (typically 10% (5% in this data)) that is omitted in the modelling and refinement process.

Figure of merit (FOM) - Accurate structures have higher figures of merit (>0.9)

Phase error- is the difference of the calculated phase from the true phase; lower figures represent more accurate structures.

6.4 Discussion

A total of 29 putatively surface located proteins in *C. difficile* have been identified as having a domain with homology to the N-acetylmuramoyl-L-alanine amidase (CwlB) of *Bacillus subtilis*. This PFam 04122 domain (cell wall binding 2 domain (CWBD)) is found in multiple repeats at either the N or C-terminus with the other end, often but not always, having a functional activity.

To date, only a fragment of the LMW S-layer subunit of *C. difficile* has been crystallised and its structure determined by X-ray crystallography (3CVZ) (Fagan *et al.* 2009). Although this structure is not the first S-layer to be crystallised (Evrard *et al.* 1999), nor the first S-layer structure to be reported (Pavkov *et al.* 2008), it is nevertheless a significant step in understanding the most important surface protein of *C. difficile*.

Given that there are only 40 structures deposited in the protein databank for *C. difficile* at the time of writing (some of which are with multiple ligands) and only one of those is a proven surface protein (3CVZ, (Fagan *et al.* 2009)), there is considerable scope to supplement the structural knowledge for *C. difficile* surface based proteins. In this chapter, selected surface proteins were expressed recombinantly, whole or as domains, in order to obtain structural information using X-ray crystallography. The three proteins chosen were:

- CD2791 an immunogenic surface located protein
- Cwp66 (Waligora *et al.* 2001) a heat shock inducible adhesin
- CD2767 a protein of unknown function but putatively containing glycosidase activity.

6.4.1 CD2791 aka Cwp2

The protein encoded by the *cd2791* gene (630 genome numbering, *cwp2* Calabi *et al.*(2001) numbering) appears to be consistently expressed and found on the cell surface of *C. difficile* during normal growth (Calabi & Fairweather 2002; Wright *et al.* 2005). Given its existence on the cell surface it is perhaps unsurprising that

CD2791 is also found in the *C. difficile* 630 spore coat (Lawley *et al.* 2009b). However, two facts bring into question the role of the CD2791 protein in *C. difficile* pathogenesis. Firstly, CD2791 (expression and protein) is absent in *C. difficile* strain 167 (a clinical isolate) indicating its non-essential nature for both cell growth and/or pathogenesis (although this strain's S-layer is also altered). Secondly, a significant number of CDI patients raise antibodies to CD2791, suggesting that antibodies raised to CD2791 are not protective (Wright *et al.* 2008), i.e. targeting this protein does not prevent disease. However, if CD2791 was non-essential, why should it be expressed so constitutively? Indeed, data in Chapter 5 suggests that knockout of CD2791 results in decreased adhesion to Caco-2 cells suggesting CD2791 is an adhesin.

6.4.1.1 Full length rCD2791

To date, recombinant expression of CD2791 has not been reported, although its promoter has been for complementation of mutants (Emerson *et al.* 2009b) (see constitutive expression above). Full length rCD2791 was expressed for two reasons; firstly any structural data obtained from crystals of rCD2791 would include key information about the CWBDs. Secondly, as there is no predicted functional activity of the N-terminus, determination of correct protein folding relied upon N-acetylmuramoyl-L-alanine amidase (amidase) activity of the two CWBDs.

The HMW SLP (which consists of three CWBDs) exhibits (amidase) activity in zymogram assay against *Micrococcus luteus* cells (Calabi *et al.* 2001). Using a similar assay it was determined that full length rCD2791 also exhibited amidase activity, confirming that other proteins CWBD also exhibit this amidase activity. The relatively low activity may have been due only to two CWBDs present at the C-terminus CD2791 (rather than 3 of the HMW SLP) or that a large proportion of the protein sample was aggregated.

Full length rCD2791 expressed well in *E. coli*, however it was apparent during purification that removal of contaminating *E. coli* protein(s) or rCD2791 truncates and aggregates was difficult. Hydrophobic interaction chromatography purified

seemingly pure rCD2791, although size exclusion chromatography revealed it was a largely aggregated mass. This data suggests that the hydrophobicity of proteins increases when there are CWBDs present. This hydrophobicity would fit well with the embedding/lower positioning in the cell wall/S-layer array.

The ability of rCD2791 to crystallise was clearly hampered by the inability to form a crystalline lattice due to the presence of this aggregated mass. Additives, e.g. detergents, chaotropes included in purification buffers, derived from their ability to remove the SLPs from the *C. difficile* cell surface, did appear to reduce aggregation, but were not successful upon scale up. Interestingly, purification of an 8 M urea SLP extract (to purify CD2791 from its native source) did not appear to purify CD2791 away from the HMW and LMW SLPs, suggesting an intimate interaction not interruptible by urea.

Together, the data suggests that it is possible to obtain soluble full length recombinant CD2791 in *E. coli.* However, the presence of two CWBDs putatively causes aggregation throughout each step of recombinant protein production. It is likely that the presence of two CWBDs of CD2791 are 'tolerated' during *E. coli* recombinant production, such that aggregation is manageable, as opposed to the HMW SLP which contains three CWBD, which required refolding (Fagan *et al.* 2009). Interestingly, expression of the 48.9 kDa conserved N-terminal of CwpV, which includes three CWBDs, was purified using a GST tag followed by SEC, did not require refolded (Emerson *et al.* 2009b). This may suggest that the CWBD, although conserved, may be sufficiently different in some surface proteins to behave differently during purification. This difference may be reflected in the attachment of the surface protein to the *C. difficile* cell wall.

6.4.1.2 N-terminal domain expression

Expression of the N-terminal region of CD2791 (rCD2791$_{27\text{-}322}$) was more successful than full length expression, whereby protein could be purified which was not aggregated. When fused to just a His$_6$ tag, rCD2791$_{27\text{-}322}$ still appeared to purify with a large amount of truncates, suggesting the rCD2791$_{27\text{-}322}$ message had

elements causing ribosomal dissociation or that rCD2791 is easily cleaved by *E. coli* proteases. The former is unlikely as the gene had putatively undergone optimisation for expression in *E. coli*. However, a C-terminal His_6 tag may have resolved these truncate purification issues. SEC analysis revealed that $rCD2791_{27\text{-}322}$ existed in solution in a monomeric form; suggesting removal of the CWBD removed the preference of CD2791 to form soluble yet aggregated masses (with itself or other *E. coli* proteins). The predicted molecular weight of $rCD2791_{27\text{-}322}$ appeared to be ~32 kDa in SDS-PAGE, however $rCD2791_{27\text{-}322}$ (post SUMO cleavage or with just a His_6 tag) migrated as a ~45 kDa species with a small percentage (a presumed truncated version) running at 32 kDa, although it is not uncommon for proteins to migrate at a different size to their predicted MW in SDS-PAGE, primarily through differences in SDS loading (Rath *et al.* 2009).

6.4.1.3 Crystallisation and modelling

Crystallisation of $rCD2791_{27\text{-}322}$ in this study started to reveal some promising hits and suggests that the N-terminal region of CD2791 maybe crystallisable given the right conditions.

There appears to be a lack of deposited structures in the PDB which have either sufficient sequence similarity or fold homology to the N-terminal domain of CD2791 for tertiary structure prediction. The proportion of β-sheets in the predicted secondary structure for $CD2791_{27\text{-}322}$ could suggest a structure similar to the heavily β-sheeted secondary structure of the Cwp66 C-terminal domain. However, it is particularly interesting that the secondary structure prediction aligns with the S-layer (SbsC) of *Geobacillus stearothermophilus* (2RA1) (Pavkov *et al.* 2008). The putative alpha-helical structure of the *C. difficile* CWBDs and the highly alpha helical cell wall binding region of SbsC may suggest a common S-layer attachment mechanism. From the data in this study, CD2791 requires crystallising for any novel fold(s) to be elucidated.

6.4.1.4 CD2791 – hypotheses

The reasons behind CD2791 having two predicted CWBDs remains elusive, but could the result of a bioinformatical error, as CD2791 may actually have three CWBD (N. Fairweather, personal communication). If CD2791 does have two CWBDs, it is possible CD2791 is not bound tightly to the cell surface, i.e. it is shed or that it can be shed if necessary, or that very C-terminal end of CD2791 (past the last CWBD) is required for another cell surface process/mechanism. There is one report of CD2791 being found in the supernatant of high toxin producing cells (Mukherjee *et al.* 2002), however, it is likely that this is from cell lysis rather than export/shedding.

The presence of CD2791 on the cell surface during normal growth suggests that it is required for some cellular process(s). As CD2791 is highly immunogenic, this suggests that the antibody responses to CD2791 are probably non-protective. Therefore, CD2791 may be a permanent decoy for the immune system and antibody responses to CD2791 do not prevent CDI. The conservation of the *cd2791* gene across strains suggests there is pressure to maintain conservation.

Given that CD2791 is putatively on the same polycistronic message as *cwp66* (Savariau-Lacomme *et al.* 2003), it would be interesting to understand how the expression of the genes on the 'operon' vary, for example, one reporter for *cd2791* (*cwp2*) was upregulated by amoxicillin, while *cwp66* is upregulated during environmental and chemical stresses (Deneve *et al.* 2008; Emerson *et al.* 2008). Increased amounts of Cwp66 would require upregulation of the entire polycistronic message, which also results in increased *cd2791* (and *cd2790*) message, unless *cd2791, cd2790* or *cwp66* are post-transcriptionally regulated.

6.4.2 Cwp66

Cwp66 was shown to be surface located particularly after heat shock with the C-terminus being presented outermost (Waligora *et al.* 2001). Consistent with this, individuals (cases or controls) appear to raise antibodies to primarily the C-terminal domain (Pechine *et al.* 2005a; Pechine *et al.* 2005b). Antibodies to Cwp66 and

recombinant Cwp66 appear to be able to prevent adhesion to Caco-2 cell, suggesting Cwp66 as an adhesin (Waligora *et al.* 2001). It is suggested that Cwp66 may be attached to the cell surface differently to the main SLPs, which can be removed with low pH treatment, as degradation of the peptidoglycan is often needed to reveal Cwp66 (Hennequin *et al.* 2001; Hennequin *et al.* 2003), even then identification of Cwp66 is not always possible (Wright *et al.* 2005).

6.4.2.1 Full length and C-terminal domain expression of Cwp66

Previous studies have determined that recombinant Cwp66 expression is when the protein is expressed as separate N- and C-terminal domains (Waligora *et al.* 2001; Pechine *et al.* 2005a). This was confirmed in this chapter, where expression of full length rCwp66 (± fusion tag) resulted in insoluble aggregates (as appears common for recombinant CWBD containing proteins).Interestingly despite Waligora *et al.* (2001) being unable to express length Cwp66, western blots of cell wall extracts detected bands at similar molecular weights (80 kDa, 50kDa and 30 kDa) to those detected by anti-His_6 during recombinant full length Cwp66 (Section 6.3.2.1, Figure 6.3.14) suggesting Cwp66 is susceptible to proteolytic cleavage. It is currently unknown if Cwp66 is post-translationally processed or is cleaved by endogenous proteases during either cell wall extraction or recombinant production.

Expression of only the C-terminal region did result in soluble protein; however, the propensity to aggregate remained, preventing successful purification. Taken together these results suggest that Cwp66 is a protein with adhesive properties, a characteristic which appears not limited to the N-terminal CWBDs.

6.4.2.2 Crystallisation and modelling

Secondary structure analysis in this chapter is consistent with data found by Waligora *et al.* (2001) whereby the predicted secondary structure of the Cwp66 C-terminal is largely β-Sheets. The arrangement of the beta sheets and the resulting function remains unknown. However, tertiary structure prediction suggested Cwp66 may utilise proteoglycans such as chondroitin sulfate (dermatan sulphate) and/or heparin, as used by *Chlamydia trachomatis* or the CS protein of *Plasmodium*

falciparum to bind host cells (Rostand & Esko 1997). Alternative tertiary structure predictions, which suggested homology to the C-terminal domain with invasin, a bacterial integrin-binding protein (1CWV) (Hamburger *et al.* 1999) and intimin, a surface bound adhesin of attachment-and-effacement bacteria (Luo *et al.* 2000) may suggest Cwp66 plays a role in bacterial adhesion, similar to type III secreted adhesins of certain Gram-negative species.

6.4.2.3 Cwp66 – hypotheses

Cwp66 has been identified as a heat shock inducible adhesin, however the co-purification of Cwp66 (and CD2767 and CD2797) with ABP labelled Cwp84 (Dang *et al.* 2010), suggests that Cwp66 may also be upregulated during surface related stress, i.e. SlpA maturation inhibition.

Antibodies to the C-terminal of Cwp66 only detected Cwp66 at the cell surface after heat shock and not in normal physiological conditions (Waligora *et al.* 2001), underlining how surface protein expression changes in response to *in vitro* conditions. It is therefore likely that surface protein expression is altered *in vivo* during CDI or during the initial colonisation stage.

The putatively tighter association of Cwp66 with the peptidoglycan and the location of the CWBD at the N-terminal of Cwp66 (rather than the C-terminal for CD2791, SlpA and Cwp84), may suggest that the presentation of the CWBDs at the N- or C-terminus dictates to what extent and how the protein interacts with a SCWP or the peptidoglycan.

Both the variability of the C-terminus of Cwp66, i.e. through selective pressure (Lemee *et al.* 2005), and that patient serum is predominantly raised to the C-terminal of Cwp66, suggest that (the C-terminal of) Cwp66 is both expressed and accessible to the immune system during infection, regardless of how putatively integrated into the cell envelope Cwp66 is.

6.4.3 CD2767

The protein encoded by the *cd2767* gene has not been investigated to date. The presence of 28 SlpA paralogs has been known about since 2001 (Calabi *et al.* 2001; Karjalainen *et al.* 2001), thus attention has largely centred on the proteins surrounding *slpA* (*cd2793*), whereas the *cd2767* gene is found downstream with a range of capsular polysaccharide biosynthesis proteins. CD2767 has not been found using any surface protein extraction techniques, however one patient was found to have antibodies to CD2767 (Wright *et al.* 2008) and upon inhibition of Cwp84, CD2767 co-purified during a pull down assay of ABP labelled Cwp84 (Dang *et al.* 2010).

6.4.3.1 Full length and C-terminal domain expression

Recombinant full length CD2767 suffered in the same way as full length expression of other surface proteins forming primarily insoluble aggregates. The addition of fusion tags partially resolved this; however purification was hampered until all the CWBDs were removed. The N-terminal domain of CD2767 ($rCD2767_{27\text{-}401}$) was extremely soluble and purified in a monomeric form in one step. It was also apparent that $rCD2767_{27\text{-}401}$ could be concentrated to very high (>100 mg/ml) concentrations. SEC analysis suggested $rCD2767_{27\text{-}401}$ was appearing as a smaller MW species than predicted (despite known % errors in SEC size estimation). This suggests that $rCD2767_{27\text{-}401}$ may be surrounded by a hydration shell masking hydrophobic residues and thus allowing increased protein concentrations (and decreased SEC MW estimation). Proteins can be preferentially hydrated (and stabilised as more compact conformations) by certain salts, amino acids and sugars, leading to their altered solubility (Jenkins 1998). Furthermore, the binding of proteins to specific salts influences solubility, through changes in protein conformation, or masking of certain amino acids involved in self-interaction (Arakawa & Timasheff 1985; Shire *et al.* 2004). Although $rCD2767_{27\text{-}401}$ exhibited increased solubility, it was apparent that stability at higher concentrations was compromised and degraded when stored at 4 °C. This process of degradation to a stable intermediate enabled the crystallisation of the LMW SLP (Fagan *et al.* 2009). Expression of stable intermediates of $rCD2767_{27\text{-}401}$ may have assisted crystal

formation, particularly if the putatively hydrated compact conformation of rCD2767$_{27\text{-}401}$ prevented the formation of a supersaturated thermodynamically meta-stable state and subsequent crystal growth.

6.4.3.2 Crystallisation and modelling

BLASTP revealed that a number of other bacteria have a proteins with very high identity (>90%) to the N-terminal domain of CD2767, suggesting that the protein domain found in the N-terminus of CD2767 (DUF187) is conserved in several species. This conservation either suggests the DUF domain is conserved for functional reasons or that it has been acquired by horizontal transfer. The identification of CD2767 on the cell surface and the reaction of patient sera to CD2767 (Wright *et al.* 2008; Dang *et al.* 2010) suggests *C. difficile* expresses CD2767 under certain conditions.

From pattern and profile searches the N-terminal domain of CD2767 matches with (trans)glycosidases of the alpha amylase superfamily. In the structural classification of proteins database (SCOP) the first seven of the TIM beta/alpha-barrel superfamilies contain a phosphate binding motif ((trans)glycosidases are number 9), thought to be of evolutionary origin (Murzin *et al.* 1995; Nagano *et al.* 2002). It is therefore interesting that IMAC purification of rCD2767$_{27\text{-}401}$ in the presence of phosphate increased the propensity of rCD2767$_{27\text{-}401}$ to form dimers and that crystals formed in the presence of phosphate (0.05 M potassium dihydrogen phosphate).

Tertiary structure predictions using EsyPred (Lambert *et al.* 2002), Phyre (Kelley & Sternberg 2009) and HHPred (Soding *et al.* 2005) largely matched with proteins in the alpha amylase superfamily (TIM barrel), while pGENThreader (Lobley *et al.* 2009) suggested CD2767$_{27\text{-}401}$ had folds similar to the beta-glycanase superfamily (TIM barrel), although some of the predicted CD2767$_{27\text{-}401}$ structures did not always correspond to the barrel region (see Figure 6.3.25). Kleiger *et al.* (2004) criticises tertiary structure servers, as a large proportion of them use a form of PSI-BLAST (pattern specific iterative BLAST). Kleiger *et al.* (2004) suggest that other methods

such as prediction of function through tertiary interactions (PFIT) or prediction of function through residues at tertiary interactions (PFRIT), may perform better at determination of glycosidase activity particularly of α/β barrel fold containing (or predicted to contain) proteins. In that, glycosidase activity can be wrongly attributed, in a larger number of cases, using PSI-BLAST than PFIT or PFRIT. However, the widespread availability of PSI-BLAST facilities makes them a simple step in determination of putative protein function. Moreover, low sequence identity is characteristic of glycosidases, which have a conserved structure rather than sequence (Nagano *et al.* 2002) and also adopt folds other than the TIM barrel (Davies & Henrissat 1995; Henrissat & Davies 1997).

Taken together, if the bioinformatic searches are correct, this suggest the N-terminal domain of CD2767 is likely to be a (trans)glycosidase, within the alpha amylase superfamily, containing a TIM (α/β barrel).

Table 6.4.1 Members of the alpha amylase superfamily from SCOP

Bacterial alpha-amylase	Maltogenic amylase, central domain	1,4-alpha-glucan branching enzyme, central domain	Oligo-1,6, glucosidase	Amylomaltase MalQ
Maltosyltransferase	Neopullulanase, central domain	Isoamylase, central domain	Isomaltulose synthase PalI	A4 beta-galactosidase
Cyclodextrin glycosyltransferase	Cyclomaltodextrinase, central domain	G4-amylase (1,4-alpha-D-glucan maltotetrahydrolase)	Amylosucrase	Bacterial beta-amylase
Animal alpha-amylase	Maltooligosyl trehalose synthase	Plant alpha-amylase	Sucrose phosphorylase	beta-Amylase
Fungal alpha-amylases	glycosyltrehalose trehalohydrolase, central domain	4-alpha-glucanotransferase	Melibiase	Pullulanase PulA

The alpha amylase superfamily belongs to the glycoside hydrolase Family 13 and contains diverse members including glycosyl transferases (Table 6.4.1). Members are retaining enzymes which act on α-glucoside containing substrates (Koshland 1953; MacGregor *et al.* 2001). Sub-specificities are defined by structural

requirements for preferred substrates or the structure of the predominant product(s).

The conserved catalytic residues of GH13 are found at the C-terminal end of the β strands with aspartic acid acting as catalytic nucleophile and glutamic acid as general acid/base, another Asp also participates in stabilising the transition state (Matsuura 2002; Nagano *et al.* 2002). These conserved residues are putatively found at the C-terminal end of strand β4 (Asp), strand β5 (glu), strand β7 (Asp) (Janecek *et al.* 2007).

Unfortunately, the structure of His.CD2767$_{27\text{-}401}$ has not been solved by molecular replacement, not unexpectedly, given the conserved structure rather than sequence of glycosidases (Nagano *et al.* 2002). Therefore, extra information is required to solve the phase problem, as an accurate enough model is not available for molecular replacement. Although, in certain cases, homology modelling can be used for molecular replacement successfully, e.g. the CaspR server (Claude *et al.* 2004). However, the identification of conditions in which crystals of His.CD2767$_{27\text{-}401}$ form, makes multiple isomorphous replacement (MIR) viable. Based on the amino acid composition of His.CD2767$_{27\text{-}401}$ Class A compounds (which non-covalently bind to the negatively charged carboxylate groups of Asp and Glu (26 x Asp and 17 x Glu) would give higher incorporation rates than Class B compounds (which bind covalently to His, Met and Cys) (Agniswamy *et al.* 2008). Heavy atom screening is considered a lengthy process both physically, e.g. crystal soaking requires upto 30 days for ethyl mercurithiosalicylate, and production of multiple heavy atom incorporated crystals requires empirical determination to ensure isomorphous incorporation. Rational approaches to heavy atom incorporation have been proposed, simplifying the process (Agniswamy *et al.* 2008). Crystals production from His.CD2767$_{27\text{-}401}$ (in MES pH 6.0) would make ideal candidates for MIR further study based on the findings of Agniswamy *et al.* (2008).

Given that there are 6 methionines in His.CD2767$_{27\text{-}401}$ (at 1, 21, 27, 117, 130, 360), it is possible that sufficient selenomethionine (SeMet) would be incorporated into

the structure, evenly spaced, for multi-wavelength anomalous dispersion (MAD) phase determination, as approximately one selenomethionine residue is required for every ~75–100 amino acids (Hendrickson *et al.* 1997; Strub *et al.* 2003). SeMet substitution is considered easier due to the ease of incorporation of recombinant proteins and may be worth pursuing with His.CD2767$_{27\text{-}401}$ given the ease of purification. Moreover, the structure of the LMW SLP (3CVZ (Fagan *et al.* 2009)) was solved by SeMet substitution and MAD, yet only has two methionines.

Due to time limitations, only preliminary crystallisation and X-ray diffraction data processing was undertaken. In terms of future prospects, it seems possible that the structure of CD2767$_{27\text{-}401}$ can be solved using MIR and/or MAD.

6.4.3.3 CD2767 – hypotheses

Why is there a protein, with homology to alpha amylase superfamily of glycosidases, presented on the cell surface only in certain conditions i.e. upon inhibition of Cwp84 (Dang *et al.* 2010) or potentially *in vivo* (Wright *et al.* 2008)?

The identification of CD2767 on the cell surface, only during certain conditions, e.g. Cwp84 inhibition or during CDI, may suggest that either: a) *cd2767*'s expression is conditional b) CD2767 is a highly labile protein or c) CD2767 is located in a different region of the cell during normal (*in vitro*) growth such that surface protein extraction techniques cannot remove it.

There are two factors that suggest that CD2767 may be involved in surface capsular biosynthesis or be involved in a coordinated surface response to stress (nutritional or otherwise). Firstly, the location of the *cd2767* gene in the *C. difficile* 630 genome next to a set of capsular polysaccharide synthesis genes, suggesting that it may be controlled with these genes rather than with other surface protein genes. Secondly, the predicted function(s) from tertiary structure prediction.

The 'highest' sequence identity of CD2767 was with the glycosyltrehalose trehalohydrolase from *Sulfolobus Solfataricus* (1EHA or 1EH9)(Feese *et al.* 2000). Glycosyltrehalose trehalohydrolase, with the aid of an intramolecular glycosyl transferase, turns soluble starch into trehalose, a non-reducing α-1,1 linked disaccharide (Feese *et al.* 2000). Trehalose is efficacious in the maintenance of cell integrity and protects against a variety of environmental injuries and nutritional limitations (Arguelles 2000). For example, during stress increasing cellular concentrations or cell surface trehalose may well protect the bacterium from damage, i.e. increase cell integrity. Tertiary structure prediction, by HHPred (Soding *et al.* 2005), also found homology to isomaltulose synthase, which isomerises sucrose to isomaltulose or trehalose (Zhang *et al.* 2003).

Three of the tertiary structure prediction servers found matches with 1,4-alpha-glucan-branching enzymes of *E. coli* and *M. tuberculosis*. In *M. tuberculosis* a glycogen branching enzyme (whose structure was recently reported 3K1D (Pal *et al.* 2010)) synthesises an α-glucan capsule which mediates non-opsonic binding to CR3 (complement receptor3 - an integrin, involved in cell-cell adhesion), blocks CD1 expression and also displays antiphagocytic properties (Pal *et al.* 2010). Therefore, CD2767 may produce (or help to produce) a capsule in *C. difficile* a) under normal conditions- however it is not found in cell surface extracts, b) in response to stress. The putative carbohydrate branching enzyme activity may be linked to secondary cell wall polymer (SCWP) construction (role during S-layer biogenesis). SCWPs are known to anchor the S-layer (non-covalently) to the cell wall (Schaffer & Messner 2005), for example, the S-layer homology domain (SLH-domain) of the *Bacillus anthracis* S-layer has been recently found to bind to a pyruvylated SCWP (Kern *et al.* 2010). Moreover, the CWBDs of *Lactobacillus acidophilus* ATCC 4356 S-layer protein shows homology to carbohydrate binding regions of *C. difficile* toxins, suggesting the involvement of carbohydrates in the cell wall binding mechanism (Smit *et al.* 2001).

Other possible roles for CD2767 include a role in the synthesis of the highly complex cell-surface teichoic-acid-like polysaccharides (PS-I and PS-II), identified by

Ganeshapillai *et al.* (2008) similar in composition to the surface carbohydrates discovered by Poxton & Cartmill (1982), composed of glucose, mannose, galactosamine and phosphate in the approximate molar proportions of 2:0.65:1:0.63 (PS-II) and glucose, glucosamine, phosphate and fatty acid (PS-I).

Surface located glycosidases have been found to play key roles in adhesion in *Streptococcus pneumonia* and *Streptococcus pyogenes*. In *S. Pneumonia*, knockout of a sortase attached surface-associated O-glycosidase decreased adhesion to human airway epithelial cells and significantly decreased its ability to colonise the upper respiratory tract (Marion *et al.* 2009). In *S. pyogenes,* a sortase bound, glycoprotein-binding and carbohydrate-degrading, pullulanase (PulA) mutant showed decreased binding to thyroglobulin (Hytonen *et al.* 2003). To date, CD2767 has not been knocked out, thus its role in *C. difficile* is still to be determined.

An intriguing concept for the surface of *C. difficile* is the possibility of cellulosomes or a modified version of cellulosomes. Cellulosomes are large extracellular enzyme complexes that break down plant cell wall polysaccharides, such as cellulose, hemicellulose and pectin into sugars (first identified in *Clostridium cellulolyticum* (Bayer & Lamed 1986; Felix & Ljungdahl 1993). The cellulosome complex consists of various kinds of enzymes (glycoside hydrolases inc. amylases, polysaccharide lyases, and carboxyl esterases) arranged around a scaffolding protein that does not exhibit catalytic activity but enables the complex to adhere to cellulose. It is postulated that the SLH domains help to bind the cellulosome to the cell surface (Doi & Kosugi 2004), for example, OlpA, OlpB and SdbA of the *C. cellulolyticum* cellulosome all contain SLH domains. The SLH domain equivalent in *C. difficile* is the CWBD and CD2767 contains three CWBDs. Furthermore, CD2797, a surface protein without CWBDs but found in the *slpA* cluster, putatively contains a dockerin type I domain which functions with a cohesion domain for complex formation in the cellulosome (Bayer *et al.* 2008). It is therefore possible that CD2767 is part of an, as yet uncharacterised cellulosome of *C. difficile*.

It is possible that when the structure of CD2767 is solved, speculation over the enzymatic mechanism/function will end. However, highly selective and specific ligand-binding sites or novel protein fold(s), may mean that a high-throughput putative ligand searching approach, similar to that used by Hermann *et al.* (2007), may be necessary to identify any substrate(s). Caution must be excercised when speculating about the structure and function of CD2767, particularly when using tertiary structure prediction. However, CD2767's co-purification with ABP-labelled Cwp84 *in vitro* by Dang *et al.* (2010) and its immunogenticity in humans (Wright *et al.* 2008), suggests CD2767 is expressed and accessible to the immune system during CDI, highlighting the need for further research to understand the role of CD2767 in *C. difficile*.

6.4.4 Summary

Recombinant expression of select surface proteins was undertaken with a view to obtaining structural information by X-ray crystallography. Three proteins were chosen, Cwp66, CD2791 (Cwp2) and CD2767. As has been found previously, removal of the CWBD repeats of *C. difficile* surface proteins, greatly aids purification to obtain pure protein. However, only one protein, CD2767, formed crystals and gave an initial data set to approximately 2 Å, although its structure was not solved by molecular replacement. However, tertiary structure prediction provided clues as to functions of some of the selected surface proteins. Overall this chapter has highlighted the need for structural data, particularly about the CWBD, to help determine how the CWBD(s) are attached to the cell surface. In conclusion, the surface of *C. difficile* is likely to be littered with proteins which play distinct roles in infection and cellular metabolism (or both), which desperately require further study.

7 General Discussion

Over the past 30 years, infection with *Clostridium difficile* has become a worldwide problem, contributing to many thousands of deaths within the developed world (Bartlett 2008b; Brazier 2008). The increased incidence throughout Europe and the US is of considerable cost, not only in terms of human suffering, but also a significant economic burden (Ghantoji *et al.* 2010).

Efforts to produce vaccines against *C. difficile* are hampered by the many aspects of CDI that remain unclear, in particular, the mechanisms by which the bacterium establishes and maintain its niche within the gut. The study of this aspect of the disease process is of paramount importance for the development of new therapeutics and prophylactics. Therefore, the proteins on the bacterial surface are at the forefront of this research need, as adherence and subsequent colonisation represents a key milestone in infection.

In this thesis, efforts to understand more about the role of surface proteins and their possible role in CDI have centred on proteins containing the putatively cell wall binding repeat domain (CWBD) at the N- or C- terminus, which displays homology to the N-acetylmuramoyl-L-alanine amidase, CwlB, of *Bacillus subtilis* (PFam04122). Some of these proteins have been implicated as playing a role in adhesion and colonisation processes, and therefore represent potential targets for therapeutic and structural research.

7.1 Entry into the Host

In order to cause disease, *C. difficile* must enter a host. The primary means for this to occur is via spores. The integrity of the spore is key to its survival and transmission, as such it can survive extremely harsh purification procedures (Lawley *et al.* 2009b) and exists on surfaces for up to 5 months (Kim *et al.* 1981). Proteomic analysis of *C. difficile* 630 spores, amongst other proteins, identified three CWBD containing proteins, SlpA, CwpV and CD2791, in the spore coat (Lawley *et al.* 2009b). However, firstly, the Cwp84 mutant produced viable spores (Kirby *et al.*

2009), suggesting SlpA maturation, or lack thereof, does not alter spore stability. Secondly, the CD2791 mutant produced viable spores, at a similar concentration to the wildtype, suggesting CD2791 does not play a role in spore viability. Thirdly, sporulation level of the CwpV mutant, although not reported to date (Emerson *et al.* 2009b; Reynolds *et al.* 2010), has been proven in our laboratory (V. Kumar, unpublished data) suggesting CwpV does not play a role in sporulation. In fact, none of the mutants produced in this thesis had altered spore viability; this suggests that the proteins investigated in this thesis do not affect spore viability or the ability to make viable spores. These results suggest, that absence of any one of the CWBD containing proteins, either on the cell surface or those found in the spore coat, do not affect *C. difficile's* ability to produce viable spores.

As the primary transmission vehicle spores, and spore specific proteins, represent a lucrative drug and vaccine target. Structural information about putative spore germination receptors (Ramirez *et al.* 2010) may enable the design inhibitors preventing germination, and preventing disease, for example, work by Howerton *et al.* (2011) has begun to identify amino acid derivates which preventing spore germination. If spore specific proteins are conserved across all *C. difficile* ribotypes, vaccination against *C. difficile* spores may largely prevent CDI in all mammalian species.

7.2 Adherence and colonisation

Given the ability of *C. difficile* surface proteins to illicit an immune response (Drudy *et al.* 2004; Pechine *et al.* 2005a; Wright *et al.* 2008), it has been suggested that increasing an immune response to these proteins, may prevent adherence and subsequently prevent CDI. Furthermore, as a large number of the *C. difficile* surface proteins can be removed using low pH or urea, they represent an easily obtainable surface antigen, for vaccine and/or immunotherapeutic investigation. Results in this thesis confirm a previous study by O'Brien *et al.* (2005), that passively administered systemic antibodies to *C. difficile* surface proteins are ultimately non-protective in the animal model, as eluted to by Pantosti *et al.* (1989). However, the decrease in colonisation observed by Pechine *et al.* (2007), suggests that inducing a more

complete mucosal protection may provide some form of colonisation resistance. Preventing colonisation before toxin production remains a key aspect to this approach.

To understand more about how *C. difficile* adheres in the gut, the identification of proteins on the bacterial surface are an obvious first line of investigation. Results in this thesis have identified four putative adhesins of *C. difficile*: Cwp66, CD2791, CD2735 and CD2795. The surface protein Cwp66 is suggested to be a heat shock inducible adhesin (Waligora *et al.* 1999), but results presented here suggest it is also plays a role in adhesion under normal conditions *in vitro*. Structural studies on CD2735, CD2795, Cwp66 and CD2791, the knock of which caused a dramatic decrease in adherence *in vitro*, may enable the design of inhibitors to prevent *C. difficile* adhesion, without requiring or needing to resort to the variable immune system.

The CWBD containing surface proteins of the *C. difficile* 027 ribotype appear conserved, a result confirmed by another group (Spigaglia *et al.* 2011a), and a result which supports the ribotyping system. The surface proteins that are not conserved (between ribotypes), e.g. SlpA, Cwp66 and CwpV, are putatively under selective pressure for immune evasion, while other, conserved, CWBD containing surface proteins, e.g. Cwp84, CD2791, are conserved to retain function, i.e. do not display genetic drift. Together with results from the knockout of *cwp66*, data suggests that Cwp66 is under strong selective pressure (as suggested by (Lemee *et al.* 2005)), because it plays a key role in adhesion. Moreover, secondary structure prediction maps variations in protein sequence, as a result of genetic variability, to loop regions, unlikely to affect the overall structure of the protein. Therefore, it is imperative that further work into a structural basis for Cwp66 ligands is performed.

In order to hasten structural studies, it appears necessary to express only the 'functional' domain(s) of surface proteins, without the accompanying CWBD. However, using this practice, one cannot obtain vital information on how the CWBD interacts with the peptidoglycan, any SCWP or other structures in the cell envelope.

The lack of sequence and particularly fold homology with deposited proteins, for example with Cwp66 and CD2971, may suggest novel folds and highlights a need for structure determination, which may or may not assist in ligand determination. The limiting factor in S-layer crystallography is the unique propensity of S-layers to form 2D arrays, rather than 3D crystalline structures. In order to obtain crystals of the CWBD (which is also likely to exhibit a propensity to form 2D arrays), it is possible some of the techniques used for membrane proteins may be employed (Wiener 2004), e.g. detergent use or any of the other methods suggested by Engelhardt & Peters (1998).

7.3 Damage to Host

A number of the mutants produced in this study had altered toxin release compared to the wildtype *C. difficile* 630ΔErm, particularly the Cwp84 and FliD mutants. However, toxin production of a Cwp84 inhibited culture (using E64 for example (Dang *et al.* 2010)) has not been reported, but if it too produced higher extracellular toxin titres (after 24 hr), care must be taken in targeting surface proteins for therapeutic interventions. It is possible, upon (chemical) inhibition of Cwp66 or CD2791, despite potentially decreasing adherence, extracellular toxin levels may increase resulting in increased mucosa damage.

The apparent link between increased toxin and decreased adhesion requires further study. Further work is also required to understand how knockout of the flagella increases adhesion but also severely increases toxin production, potentially causing a more virulent strain (Baban *et al.* 2010). Recent work suggesting that activation of TLR5 with *Salmonella typhimurium* flagellin appears to protect mice from CDI (Jarchum *et al.* 2011), further complicates the role of flagella and flagellins in both colonisation and CDI.

This apparent inverse relationship between toxin and adherence has been described in a histone-like nucleoid structuring regulatory gene (*hns*) mutant of *E. coli* 091:H21 strain B2F1 (Scott *et al.* 2003). This mutant showed decreased adherence to T84 human colonic epithelial cells and an increase in the production

of hemolysin. H-NS proteins putatively play a role regulating gene expression, including regulation of adhesins (Beltrametti *et al.* 1999), although they have not been found in Gram-positive bacteria (Tendeng & Bertin 2003).

It is possible that knockout of certain surface proteins, alters cell integrity resulting in increased bacterial stress (similar to nutrient limiting conditions) causing changes in toxin levels, or that the physical loss of that surface protein results in a decrease in cellular integrity and results in the release of intracellular toxin, i.e. toxin leakage hypothesis. Under this latter hypothesis, any surface protein mutant with toxin levels comparable to the wild-type may not have sufficiently altered cell integrity, detectable by toxin changes, however, confirms that the regulation of virulence factors is complex and requires further work.

7.4 Release and spread

The ability of *C. difficile* to transmit relies on the ability to produce viable spores and, as aforementioned, all of the surface proteins mutants created in this thesis produced viable spores. Although, it has been found that a mutation in a protein synthesised during sporulation and incorporated into the spores, affects morphology but not viability (Mallozzi *et al.* 2008). Therefore, it would be interesting to determine if surface protein mutant spores are affected morphologically, particularly for *C. difficile* Cwp84, CD2791 or CwpV mutants given their connection with the spore coat.

7.5 Therapeutics

The goal of a *C. difficile* therapy, particularly a vaccine, is to prevent CDI occurring, although, as suggested by computer modelling, targeting *C. difficile* recurrence may also be lucrative (Lee *et al.* 2010a). However, there is understandably caution over whom to vaccinate and even if vaccination is effective, whether it is likely to fail in those whose immune system becomes compromised.

There are several approaches to the treatment of CDI, including a plethora of 'new' antibiotics, some of which are likely to be more successful at treating CDI than

others (Bartlett 2009). Use of an *in vitro* gut model has been shown to be remarkably successful in identifying therapeutics which have gone on to fail in clinical trial (Baines *et al.* 2009) and may therefore represent a 'simple' test for appropriate forthcoming anti-*C. difficile* agents. There are also other 'out of the box' therapies, briefly discussed in Chapter 4, including chemical toxin neutralisation, restoration of the protective gut flora and immunotherapy (active and passive) (Gerding & Johnson 2010). Gerding and Johnson (2010) point out how several of these 'non-traditional' therapeutics appear promising, which together with the rapidly growing number of bacterial species resistant to multiple antibiotics, e.g. *Mycobacterium tuberculosis* (Caminero *et al.* 2010), puts real impetus on finding new ways to treat infection.

Techniques to target virulence, i.e. the ability of the bacterium to cause disease, rather than simply the targeting of bacterial existence (as antibiotics currently do), has the putative advantage that it may not impart as high selective pressure for resistance than current antibiotics (Clatworthy *et al.* 2007; Cegelski *et al.* 2008).

In light of the results of this thesis, what is the future, or what is required for a future into a surface, i.e. anti-adhesion/colonisation, therapy for *C. difficile*?

Presentation of an immature *C. difficile* S-layer at the cell surface results in decreased cell integrity (Kirby *et al.* 2009; Dang *et al.* 2010), therefore targeting S-layer biogenesis to prevent CDI, as suggested by Dang *et al.* (2010), may provide a novel way to prevent CDI via a mechanism not dissimilar to β-lactam antibiotics. Therefore, there may be an opportunity to develop new therapies based on the aberration of cell surface (S-layer) construction. As the SLPs of *C. difficile* can be removed via chemical methods, e.g. low pH, LiCl, EDTA or denaturing agents, and as demonstrated by Takumi *et al.* (1991), the loss of the SLPs results in an inability to adhere to the gut, development of a chemical therapy to dissociate the SLPs, i.e. prevent adherence, is feasible (although the author would not suggest an 8 M urea enema to 'resolve' CDI). However, in order to even contemplate development of a novel CWBD neutralising compound, research is required into understanding how

the CWBD interacts with the peptidoglycan, a SCWP or other structure(s) in the cell envelope, structural research would be key in this respect.

The apparent inability of passively administered anti-SLP IgG to prevent CDI in the animal model (Chapter 4 and O'Brien *et al.* (2005)), may point to differences in the presentation of surface proteins *in vitro* to *in vivo*, or an inability of the antibodies to access the gut when they are required to prevent adhesion. Development of anti-surface protein (immuno)therapeutics must therefore ensure that the *in vitro* antigen, matches the *in vivo* form. Having said that, patients who have CDI are presented with *C. difficile* surface antigens in their 'correct' form, but whether (humoral) responses to the surface proteins are protective requires further study (Drudy *et al.* 2004; Pechine *et al.* 2007; Sanchez-Hurtado *et al.* 2008; Wright *et al.* 2008). Therefore, the key to anti-colonisation immunotherapy may be in generating the correct type of protective immune response. Lessons learnt from *C. difficile* toxin therapy research, has revealed that the native toxins raise a large immune response, but it is not protective, i.e. not cytotoxin neutralising. Anti-SLP antibodies, may, therefore cause a similar non-protective response, except only act on one part of the infection cycle e.g. adhesion. To resolve the toxin antibody paradox, toxins are toxoided, i.e. chemically modified with formaldehyde (Salnikova *et al.* 2008). Therefore, to increase the ability of the SLPs to raise a protective immune response, it is possible they require 'fixing', possibly preventing the presentation of highly immunoreactive, but non-protecting antibodies. The protein CD2791 may be particularly good at this at raising non-protecting antibodies, as all CDI patients were found to raise antibodies to CD2791 (Wright *et al.* 2008).

The conservation of the large majority of the SLPs, and the variability of relatively few, suggests that different surface proteins have different selective pressures put upon them, e.g. variation for immune evasion (SlpA and Cwp66 (Lemee *et al.* 2005)) or conservation of action (Cwp84). Therefore, developing on results in this thesis, knockout of all 28 *slpA* paralogs, and assessment of their virulence *in vivo*, may be necessary to understand if any of the paralogs have roles in *C. difficile* pathogenesis.

7.6 Identification of factors necessary for infection

C. difficile in vitro gene expression is modified by environmental factors (Waligora *et al.* 1999; Karlsson *et al.* 2003; Deneve *et al.* 2008; Emerson *et al.* 2008; Karlsson *et al.* 2008; Deneve *et al.* 2009a). However, it is almost impossible to account for, *in vitro*, the complex milieu of the gut, host-cell interactions and the dynamic environmental stimuli associated with the infection process *in vivo* (Acheson & Luccioli 2004; O'Hara & Shanahan 2006; Kim *et al.* 2010).

However, there are a number of genetic techniques which have yet to be applied to *C. difficile* pathogenesis studies, that may help elucidate key genes/proteins associated with infection *in vivo.* These are: differential fluorescence induction (DFI), *in vivo* expression technology (IVET), signature tagged mutagenesis (STM) and *in vivo*-induced antigen technology (IVIAT) (reviewed in Hautefort *et al.* (2002)).

7.6.1 Differential Fluorescence Induction (DFI) (Bumann & Valdivia 2007)

DFI involves the creation of a library of random genomic fragments, fused to a promotorless *gfp* gene. The fragments are cloned into a plasmid and entered into the organism of interest. If a fragment contains a promoter, active during certain stimuli, e.g. *in vivo* or infecting mammalian cells, GFP will be expressed and fluorescent cells can be isolated by FACS. Comparison of the fragment with the genome sequence, allows the identification of genes whose products may be associated with the response of the organism to the condition assessed. Therefore, application of DFI to *C. difficile* requires a plasmid stable in *C. difficile*, the pMTL80000 series of vectors (Heap *et al.* 2009b), could fulfil this need. However, DFI is subject to criticism, such as, use of a fluorophore and an inability to measure post-transcriptionally regulated genes (Rediers *et al.* 2005).

7.6.2 *In vivo* expression technology (IVET) (Angelichio & Camilli 2002)

IVET, similar to DFI, uses 'promoter trapping' and positive selection in that it selects for genes that are expressed *in vivo*. Classical IVET requires an essential growth factor (*egf*) mutant, and a plasmid containing a promotorless copy of the *egf* gene and a reporter gene. Random bacterial DNA is cloned into the plasmid which is then

integrated in the chromosome of the *egf* mutant strain. The Δegf mutants, with insertions, are then infected into the appropriate animal model. When the promoter is expressed, the EGF is returned, i.e. the mutant is complemented, thus the strain causes disease and can be isolated from the target organ(s). Recovered bacteria are plated and reporter negative clones are analysed. There have been modifications to IVET over time including use of antibiotic selection rather than an auxotroph, dual reporter system or inheritable antibiotic-sensitive phenotype. Use of IVET in *C. difficile* would require an auxotroph, presumably created with the ClosTron system and a plasmid able to homologously recombine in *C. difficile*, potentially as used in Lyras *et al.* (2009), as stability is not required. IVET, if used with antibiotic selection, amongst other limitations, is likely to result in transcription of genes associated with exposure of the organism to that antibiotic, e.g. *C. difficile* to clindamycin (Deneve *et al.* 2009a), rather than as a result of the infection process (Rediers *et al.* 2005).

7.6.3 Signature tagged mutagenesis (STM) (Hensel *et al.* 1995; Autret & Charbit 2005)

As opposed to IVET or DFI, STM uses negative selection to identify genes required for infection *in vivo*. A library of random mutants is created with individually tagged transposons. The viable mutants are pooled then assessed for their ability to cause disease in the appropriate animal model. If a mutant cannot cause infection, then that mutant's tag will not be detected among the recovered bacteria. Therefore, STM identifies genes important for virulence and survival *in vivo*, rather than just genes that are switched on (as IVET or DFI does). Therefore, for STM there is a requirement for random insertional mutagenesis, the recently reported transposon systems by Hussain *et al.* (2010) and Cartman & Minton (2010) are prime candidates for the development of STM in *C. difficile*. Many of the limitations of STM revolve around the use of a mutant pool, true randomness of the transposon and hybridisation issues.

7.6.4 *In vivo* induced antigen technology (IVIAT) (Handfield *et al.* 2000)

Perhaps the 'easiest' of the techniques to identify factors expressed during infection, in terms of applicability to *C. difficile*, i.e. without genetic modification, is that of IVIAT. IVIAT uses convalescent sera which are exhaustively absorbed against bacteria grown *in vitro*, leaving only antibodies against antigens that are expressed *in vivo*. The adsorbed serum is then used to probe an *E. coli* inducible protein expression library, reacting clones are then sequenced and the protein(s) identified. Additional confirmation of the expression *in vivo* can be performed by, for example, immunohistochemistry, using fluorescent-labelled antibodies to screen biological specimens. IVIAT has obvious advantages in that it does not require animal models, or any genetic manipulation of the disease causing organism. However, one of the major flaws of IVIAT is that it can identify proteins that are good immunogens *in vivo*, but not necessarily required for virulence. As demonstrated for *C. difficile* toxins, a good immune response does not necessarily equal a good neutralising titre (Giannasca *et al.* 1999). IVIAT could be considered an extension of work by Wright *et al.* (2008), where rather than proteomic analysis of defined surface antigens using patient serum, a genomic library would be screened. IVIAT has been successfully used to identify several surface proteins of *B. anthracis* involved in infection (Rollins *et al.* 2008), highlighting IVIATs applicability in identification of (surface based) factors associated with *in vivo* infection.

7.7 Summary

C. difficile is a serious problem, not only in the healthcare setting but increasingly in the community. The results presented in this thesis have implications in different steps of *C. difficile* pathogenesis, and in furthering *C. difficile* knowledge into the role of surface proteins in CDI. In particular, only the *C. difficile* Cwp84 surface protein mutant had altered sporulation ability but not viability, suggesting vegetative cell surface proteins are not involved in spore stability and viability. It is also apparent that *C. difficile* has a highly integrated metabolism and regulatory system, as surface proteins and toxin production are connected in some way. New treatments for *C. difficile* infection are required that both prevent and treat CDI, but

also prevent relapses. Immunotherapy targeting bacterial surface proteins, from the circulation to prevent colonisation, did not prevent CDI in the hamsters model, suggesting other (potentially immunotherapeutic) anti-colonisation approaches be sought. One of the key aims of *C. difficile* research will be in the identification of (surface) proteins involved in infection. In this respect, newly developed genetic tools make identification of these key virulence factors a real possibility.

8 Conclusions

In order to understand more about how *C. difficile* colonises the gut, research was undertaken into the surface proteins of *C. difficile,* in particular, proteins which contain a PFam04122 domain, which putatively mediates their attachment to the cell surface.

Using a recently developed knockout system, a number of novel surface protein knockouts were made. A total of eight surface protein encoding genes were knocked out (*cd1036*, *cd2735*, *cd2784*, *cd2791*, *cwp66*, *cwp84*, *cd2795* and *fliD*) and several of these were found to be key in bacterial surface physiology and adhesion. Of key importance was the knockout of Cwp84, which results in the presentation of immature SlpA (S-layer precursor) on the cell surface. Efforts to obtain structural information about the surface proteins CD2791, Cwp66 and CD2767, by X-ray crystallography were hampered by the PFam04122 repeats, as their removal significantly aided purification. However, an initial dataset to 2 Å was obtained from a crystal of the N-terminal domain of CD2767. Efforts to understand any surface based reason for increased virulence of the 'epidemic' 027 ribotype, revealed that that Pfam04122 containing surface proteins are largely similar genetically and that passively administered antibodies, raised to low pH extracted 027 ribotype surface proteins, are not protective in hamsters.

8.1 Summary of contributions

Using a knockout approach, it was discovered that the surface-located cysteine protease Cwp84 cleaves SlpA and presentation of an immature S-layer on the bacterial cell surface severely affects bacterial physiology. CD2735, Cwp66, CD2791, and CD2795 mutants had decreased adhesion *in vitro* and, particularly for FliD and Cwp84 mutants, displayed higher toxin A production after 24 hr. Inactivation of peptidoglycan hydrolase domain containing surface protein CD2784 slows bacterial growth and potentially increases adhesion, while knockout of CD1036, which has a similar domain structure to CD2784, has no observable affect on *C. difficile* function *in vitro*. All surface protein mutants produced viable spores.

Genetically the Pfam04122 containing surface proteins of *C. difficile* are largely conserved, apart from *slpA*, *cwp66* and *cwpV*. Moreover, passively administered (systemic) antibodies to an epidemic 027 ribotype strain (R20291) do not protect hamsters from CDI.

To readily obtain pure protein for crystallisation, the PFam04122 repeats needed to be removed, i.e. the protein expressed without these repeats. Once this was accomplished, crystals of the N-terminal DUF187 domain of CD2767 were grown and diffracted to 2 Å, although the structure was not solved by molecular replacement due to a lack of similar structures in the protein databank.

8.2 Future work

As outlined throughout this thesis, there is still much work to be done regarding understanding the mechanisms by which *C. difficile* adheres to the gut.

While Cwp84 mediated cleavage of SlpA has been conclusively proven, further work is required to fully understand S-layer biogenesis, particularly as post-translational cleavage of the S-layer is unique to *C. difficile*. Understanding specifics of the cleavage mechanism would be significantly aided by structural information for both Cwp84 and full length SlpA, for example, in identifying the likely conformational change of SlpA after cleavage. Identification of the location of cleavage, i.e. where on or in the cell wall cleavage occurs, and the point of Cwp84 (auto)maturation also requires further work. Underlying this research is the requirement to understand how the cell wall binding domain (Pfam 04122) (CWBD), present in 28 other putative *C. difficile* surface proteins, is connected to the cell envelope, secondary cell wall polymers (SWCPs) or peptidoglycan via a non-covalent mechanism. Many other bacterial S-layers cell wall binding domains have been found to use SWCPs, although such a polymer has yet to be identified in *C. difficile*.

To date, not all of the 28 SlpA paralogs have been found on the cell surface or even their expression proven. It is apparent that further work is required to understand if, when, and why these proteins are or are not expressed and their role in *C. difficile* physiology and pathogenesis. Investigations into the connection (or apparent connection) between surface proteins and other systems, e.g. toxin production, are required, particularly as toxins are the major virulence factor of *C. difficile*.

Structural studies, X-ray, NMR or electron based, are key in furthering knowledge into S-layer biogenesis and the myriad of proteins found on the cell surface, particularly those found to be important *in vivo*. In this respect, the development of anti-colonisation inhibitors or vaccines would be significantly aided by structural data. Moreover, structural information may provide physical links to functionality, particularly for those surface proteins for which no function can be assigned at present.

The role of the immune system in *C. difficile* infection is one area only touched on in this thesis. However, recent data suggests that both colonisation and development of CDI can be controlled by management of the immune system (Jarchum *et al.* 2011). Research into anti-colonisation therapy may need to look beyond serum based responses to a mucosal disease. Research also needs to clarify at what point during infection adhesion, colonisation and toxin production occur, and where the immune system plays a role in those processes.

The ubiquitous nature of *C. difficile* spores in the environment means acquisition is highly probable, for humans or animals. Research into the *C. difficile* spore structure and spore germination is required to fully understand the *C. difficile* delivery vehicle. A therapy aimed at resolving CDI by preventing the sporulation pathway may decrease a) the chances of patient relapse, b) transmission and contamination of the environment.

The role of genetic manipulation of *C. difficile* will provide much needed information about all aspects of *C. difficile* pathogenesis and physiology. However, a limitation of the ClosTron system is that only one gene is knocked out at a time and theoretically requires each mutant to be assessed for its ability to cause CDI, *in vivo*, to identify factors key for infection. Approaches outlined for the identification of multiple genes expressed *in vivo,* e.g. STM, IVET, may be key in this respect and require investigation.

Publications

Kirby, J. M., Ahern, H., Roberts, A. K., Kumar, V., Freeman, Z., Acharya, K. R. and Shone, C. C. (2009). Cwp84, a surface-associated cysteine protease, plays a role in the maturation of the surface layer of Clostridium difficile. *J Biol Chem*, **284**, 34666-34673.

References

Abad, M. C., Binderup, K., Rios-Steiner, J., Arni, R. K., Preiss, J. and Geiger, J. H. (2002). The X-ray crystallographic structure of Escherichia coli branching enzyme. *J Biol Chem*, **277**, 42164-42170.

Aboudola, S., Kotloff, K. L., Kyne, L., Warny, M., Kelly, E. C., Sougioultzis, S., Giannasca, P. J., Monath, T. P. and Kelly, C. P. (2003). Clostridium difficile vaccine and serum immunoglobulin G antibody response to toxin A. *Infect Immun*, **71**, 1608-1610.

Abougergi, M. S., Broor, A., Cui, W. and Jaar, B. G. (2010). Intravenous immunoglobulin for the treatment of severe Clostridium difficile colitis: an observational study and review of the literature. *J Hosp Med*, **5**, E1-9.

Acheson, D. W. and Luccioli, S. (2004). Microbial-gut interactions in health and disease. Mucosal immune responses. *Best Pract Res Clin Gastroenterol*, **18**, 387-404.

Ackermann, G., Tang, Y. J., Henderson, J. P., Rodloff, A. C., Silva, J., Jr. and Cohen, S. H. (2001). Electroporation of DNA sequences from the pathogenicity locus (PaLoc) of toxigenic Clostridium difficile into a non-toxigenic strain. *Mol Cell Probes*, **15**, 301-306.

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. and Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*, **66**, 213-221.

Aghajari, N., Feller, G., Gerday, C. and Haser, R. (1998). Crystal structures of the psychrophilic alpha-amylase from Alteromonas haloplanctis in its native form and complexed with an inhibitor. *Protein Sci*, **7**, 564-572.

Agniswamy, J., Joyce, M. G., Hammer, C. H. and Sun, P. D. (2008). Towards a rational approach for heavy-atom derivative screening in protein crystallography. *Acta Crystallogr D Biol Crystallogr*, **64**, 354-367.

Akerlund, T., Persson, I., Unemo, M., Noren, T., Svenungsson, B., Wullt, M. and Burman, L. G. (2008). Increased sporulation rate of epidemic Clostridium difficile Type 027/NAP1. *J Clin Microbiol*, **46**, 1530-1533.

Al-Nassir, W. N., Sethi, A. K., Li, Y., Pultz, M. J., Riggs, M. M. and Donskey, C. J. (2008). Both oral metronidazole and oral vancomycin promote persistent overgrowth of vancomycin-resistant enterococci during treatment of Clostridium difficile-associated disease. *Antimicrob Agents Chemother*, **52**, 2403-2406.

Albesa-Jove, D., Bertrand, T., Carpenter, E. P., Swain, G. V., Lim, J., Zhang, J., Haire, L. F., Vasisht, N., Braun, V., Lange, A., von Eichel-Streiber, C., Svergun, D. I., Fairweather, N. F. and Brown, K. A. (2010). Four distinct structural domains in Clostridium difficile toxin B visualized using SAXS. *J Mol Biol*, **396**, 1260-1270.

Allen, S. D., Emery, C. L. and Lyerly, D. M. (2003). Clostridium. *Manual of clinical microbiology*. Murray, P. R., Baron, E. J., Jorgensen, J. H., Pfaller, M. A. and Yolken, R. H., American Society for Microbiology, Washington, D.C**:** 835–856.

Angelichio, M. J. and Camilli, A. (2002). In vivo expression technology. *Infect Immun*, **70**, 6518-6523.

Arakawa, T. and Timasheff, S. N. (1985). Theory of protein solubility. *Methods Enzymol*, **114**, 49-77.

Arguelles, J. C. (2000). Physiological roles of trehalose in bacteria and yeasts: a comparative analysis. *Arch Microbiol*, **174**, 217-224.

Arnold, K., Bordoli, L., Kopp, J. and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, **22**, 195-201.

Aronsson, B., Granstrom, M., Mollby, R. and Nord, C. E. (1985). Serum antibody response to Clostridium difficile toxins in patients with Clostridium difficile diarrhoea. *Infection*, **13**, 97-101.

Arora, S. K., Dasgupta, N., Lory, S. and Ramphal, R. (2000). Identification of two distinct types of flagellar cap proteins, FliD, in Pseudomonas aeruginosa. *Infect Immun*, **68**, 1474-1479.

Ausiello, C. M., Cerquetti, M., Fedele, G., Spensieri, F., Palazzo, R., Nasso, M., Frezza, S. and Mastrantonio, P. (2006). Surface layer proteins from Clostridium difficile induce inflammatory and regulatory cytokines in human monocytes and dendritic cells. *Microbes Infect*, **8**, 2640-2646.

Autret, N. and Charbit, A. (2005). Lessons from signature-tagged mutagenesis on the infectious mechanisms of pathogenic bacteria. *FEMS Microbiol Rev*, **29**, 703-717.

Baban, S., Kuehne, S. A., Hardie, K., Barketi, A., Kansau, I., Collignon, A. and Minton, N. (2010). Role of Flagella in aspects of Clostridium difficile virulence. 3rd International Clostridium difficile symposium, Bled, Slovenia.

Babcock, G. J., Broering, T. J., Hernandez, H. J., Mandell, R. B., Donahue, K., Boatright, N., Stack, A. M., Lowy, I., Graziano, R., Molrine, D., Ambrosino, D. M. and Thomas, W. D., Jr. (2006). Human monoclonal antibodies directed against toxins A and B prevent Clostridium difficile-induced mortality in hamsters. *Infect Immun*, **74**, 6339-6347.

Backert, S., Tegtmeyer, N. and Selbach, M. (2010). The versatility of Helicobacter pylori CagA effector protein functions: The master key hypothesis. *Helicobacter*, **15**, 163-176.

Bahl, H., Scholz, H., Bayan, N., Chami, M., Leblon, G., Gulik-Krzywicki, T., Shechter, E., Fouet, A., Mesnage, S., Tosi-Couture, E., Gounon, P., Mock, M., Conway de Macario, E., Macario, A. J., Fernandez-Herrero, L. A., Olabarria, G., Berenguer, J., Blaser, M. J., Kuen, B., Lubitz, W., Sara, M., Pouwels, P. H., Kolen, C. P., Boot, H. J. and Resch, S. (1997). Molecular biology of S-layers. *FEMS Microbiol Rev*, **20**, 47-98.

Baines, S. D., Freeman, J. and Wilcox, M. H. (2009). Tolevamer is not efficacious in the neutralization of cytotoxin in a human gut model of Clostridium difficile infection. *Antimicrob Agents Chemother*, **53**, 2202-2204.

Bakken, J. S. (2009). Fecal bacteriotherapy for recurrent Clostridium difficile infection. *Anaerobe*, **15**, 285-289.

Baldwin, M. A. (2004). Protein identification by mass spectrometry: issues to be considered. *Mol Cell Proteomics*, **3**, 1-9.

Barrett, A. J. and Rawlings, N. D. (2001). Evolutionary lines of cysteine peptidases. *Biol Chem*, **382**, 727-733.

Bartlett, J. G. (2002). Clinical practice. Antibiotic-associated diarrhea. *N Engl J Med*, **346**, 334-339.

Bartlett, J. G. (2006). Narrative review: the new epidemic of Clostridium difficile-associated enteric disease. *Ann Intern Med*, **145**, 758-764.

Bartlett, J. G. (2008a). The case for vancomycin as the preferred drug for treatment of Clostridium difficile infection. *Clin Infect Dis*, **46**, 1489-1492.

Bartlett, J. G. (2008b). Historical perspectives on studies of Clostridium difficile and C. difficile infection. *Clin Infect Dis*, **46 Suppl 1**, S4-11.

Bartlett, J. G. (2009). New antimicrobial agents for patients with Clostridium difficile infections. *Curr Infect Dis Rep*, **11**, 21-28.

Bartlett, J. G., Onderdonk, A. B., Cisneros, R. L. and Kasper, D. L. (1977). Clindamycin-associated colitis due to a toxin-producing species of Clostridium in hamsters. *J Infect Dis*, **136**, 701-705.

Bauer, M. P., Notermans, D. W., van Benthem, B. H., Brazier, J. S., Wilcox, M. H., Rupnik, M., Monnet, D. L., van Dissel, J. T. and Kuijper, E. J. (2011). Clostridium difficile infection in Europe: a hospital-based survey. *Lancet*, **377**, 63-73.

Bayer, E. A. and Lamed, R. (1986). Ultrastructure of the cell surface cellulosome of Clostridium thermocellum and its interaction with cellulose. *J Bacteriol*, **167**, 828-836.

Bayer, E. A., Lamed, R., White, B. A. and Flint, H. J. (2008). From cellulosomes to cellulosomics. *Chem Rec*, **8**, 364-377.

Beltrametti, F., Kresse, A. U. and Guzman, C. A. (1999). Transcriptional regulation of the esp genes of enterohemorrhagic Escherichia coli. *J Bacteriol*, **181**, 3409-3418.

Berg, J. M., Tymoczko, J. L. and Stryer, L. (2002). Biochemistry. New York, W. H. Freeman.

Best, E. L., Fawley, W. N., Parnell, P. and Wilcox, M. H. (2010). The potential for airborne dispersal of Clostridium difficile from symptomatic patients. *Clin Infect Dis*, **50**, 1450-1457.

Beveridge, T. J. (1990). Mechanism of gram variability in select bacteria. *J Bacteriol*, **172**, 1609-1620.

Blair, D. E., Schuttelkopf, A. W., MacRae, J. I. and van Aalten, D. M. (2005). Structure and metal-dependent mechanism of peptidoglycan deacetylase, a streptococcal virulence factor. *Proc Natl Acad Sci U S A*, **102**, 15429-15434.

Blaser, M. J. and Pei, Z. (1993). Pathogenesis of Campylobacter fetus infections: critical role of high-molecular-weight S-layer proteins in virulence. *J Infect Dis*, **167**, 372-377.

Bobulsky, G. S., Al-Nassir, W. N., Riggs, M. M., Sethi, A. K. and Donskey, C. J. (2008). Clostridium difficile skin contamination in patients with C. difficile-associated disease. *Clin Infect Dis*, **46**, 447-450.

Bondos, S. E. and Bicknell, A. (2003). Detection and prevention of protein aggregation before, during, and after purification. *Anal Biochem*, **316**, 223-231.

Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J. and Schwede, T. (2009). Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protoc*, **4**, 1-13.

Bork, P., Holm, L. and Sander, C. (1994). The immunoglobulin fold. Structural classification, sequence patterns and common core. *J Mol Biol*, **242**, 309-320.

Borriello, S. P. (1990). The influence of the normal flora on Clostridium difficile colonisation of the gut. *Ann Med*, **22**, 61-67.

Borriello, S. P. and Barclay, F. E. (1985). Protection of hamsters against Clostridium difficile ileocaecitis by prior colonisation with non-pathogenic strains. *J Med Microbiol*, **19**, 339-350.

Borriello, S. P., Welch, A. R., Barclay, F. E. and Davies, H. A. (1988). Mucosal association by Clostridium difficile in the hamster gastrointestinal tract. *J Med Microbiol*, **25**, 191-196.

Branden, C. and Tooze, J. (1999). Introduction to protein structure. New York, Garland Pub.

Brazier, J. S. (1998). The epidemiology and typing of Clostridium difficile. *J Antimicrob Chemother*, **41 Suppl C**, 47-57.

Brazier, J. S. (2008). Clostridium difficile: from obscurity to superbug. *Br J Biomed Sci*, **65**, 39-44.

Brazier, J. S., Raybould, R., Patel, B., Duckworth, G., Pearson, A., Charlett, A. and Duerden, B. I. (2008). Distribution and antimicrobial susceptibility patterns of Clostridium difficile PCR ribotypes in English hospitals, 2007-08. *Euro Surveill*, **13**.

Breitwieser, A., Gruber, K. and Sleytr, U. B. (1992). Evidence for an S-layer protein pool in the peptidoglycan of Bacillus stearothermophilus. *J Bacteriol*, **174**, 8008-8015.

Brun, P., Scarpa, M., Grillo, A., Palu, G., Mengoli, C., Zecconi, A., Spigaglia, P., Mastrantonio, P. and Castagliuolo, I. (2008). Clostridium difficile TxAC314 and SLP-36kDa enhance

the immune response toward a co-administered antigen. *J Med Microbiol*, **57**, 725-731.

Brunger, A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472-475.

Buist, G., Steen, A., Kok, J. and Kuipers, O. P. (2008). LysM, a widely distributed protein motif for binding to (peptido)glycans. *Mol Microbiol*, **68**, 838-847.

Bumann, D. and Valdivia, R. H. (2007). Identification of host-induced pathogen genes by differential fluorescence induction reporter systems. *Nat Protoc*, **2**, 770-777.

Burden, R. E., Snoddy, P., Buick, R. J., Johnston, J. A., Walker, B. and Scott, C. J. (2008). Recombinant cathepsin S propeptide attenuates cell invasion by inhibition of cathepsin L-like proteases in tumor microenvironment. *Mol Cancer Ther*, **7**, 538-547.

Burns, D. A., Heap, J. T. and Minton, N. P. (2010a). The diverse sporulation characteristics of Clostridium difficile clinical isolates are not associated with type. *Anaerobe*, **16**, 618-622.

Burns, D. A., Heap, J. T. and Minton, N. P. (2010b). SleC is essential for germination of Clostridium difficile spores in nutrient-rich medium supplemented with the bile salt taurocholate. *J Bacteriol*, **192**, 657-664.

Burns, K., Morris-Downes, M., Fawley, W. N., Smyth, E., Wilcox, M. H. and Fitzpatrick, F. (2010c). Infection due to C. difficile ribotype 078: first report of cases in the Republic of Ireland. *J Hosp Infect*, **75**, 287-291.

Butt, T. R., Edavettal, S. C., Hall, J. P. and Mattern, M. R. (2005). SUMO fusion technology for difficult-to-express proteins. *Protein Expr Purif*, **43**, 1-9.

Calabi, E., Calabi, F., Phillips, A. D. and Fairweather, N. F. (2002). Binding of Clostridium difficile surface layer proteins to gastrointestinal tissues. *Infect Immun*, **70**, 5770-5778.

Calabi, E. and Fairweather, N. (2002). Patterns of sequence conservation in the S-Layer proteins and related sequences in Clostridium difficile. *J Bacteriol*, **184**, 3886-3897.

Calabi, E., Ward, S., Wren, B., Paxton, T., Panico, M., Morris, H., Dell, A., Dougan, G. and Fairweather, N. (2001). Molecular characterization of the surface layer proteins from Clostridium difficile. *Mol Microbiol*, **40**, 1187-1199.

Camiade, E., Peltier, J., Bourgeois, I., Couture-Tosi, E., Courtin, P., Antunes, A., Chapot-Chartier, M. P., Dupuy, B. and Pons, J. L. (2010). Characterization of Acp, a peptidoglycan hydrolase of Clostridium perfringens with N-acetylglucosaminidase activity that is implicated in cell separation and stress-induced autolysis. *J Bacteriol*, **192**, 2373-2384.

Caminero, J. A., Sotgiu, G., Zumla, A. and Migliori, G. B. (2010). Best drug treatment for multidrug-resistant and extensively drug-resistant tuberculosis. *Lancet Infect Dis*, **10**, 621-629.

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res*, **37**, D233-238.

Cao, M. and Helmann, J. D. (2004). The Bacillus subtilis extracytoplasmic-function sigmaX factor regulates modification of the cell envelope and resistance to cationic antimicrobial peptides. *J Bacteriol*, **186**, 1136-1146.

Cartman, S. T., Heap, J. T., Kuehne, S. A., Cockayne, A. and Minton, N. P. (2010). The emergence of 'hypervirulence' in Clostridium difficile. *Int J Med Microbiol*, **300**, 387-395.

Cartman, S. T. and Minton, N. P. (2010). A mariner-based transposon system for in vivo random mutagenesis of Clostridium difficile. *Appl Environ Microbiol*, **76**, 1103-1109.

CCP4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*, **50**, 760-763.

Cegelski, L., Marshall, G. R., Eldridge, G. R. and Hultgren, S. J. (2008). The biology and future prospects of antivirulence therapies. *Nat Rev Microbiol*, **6**, 17-27.

Cerquetti, M., Molinari, A., Sebastianelli, A., Diociaiuti, M., Petruzzelli, R., Capo, C. and Mastrantonio, P. (2000). Characterization of surface layer proteins from different Clostridium difficile clinical isolates. *Microb Pathog*, **28**, 363-372.

Cerquetti, M., Pantosti, A., Stefanelli, P. and Mastrantonio, P. (1992). Purification and characterization of an immunodominant 36 kDa antigen present on the cell surface of Clostridium difficile. *Microb Pathog*, **13**, 271-279.

Cerquetti, M., Serafino, A., Sebastianelli, A. and Mastrantonio, P. (2002). Binding of Clostridium difficile to Caco-2 epithelial cell line and to extracellular matrix proteins. *FEMS Immunol Med Microbiol*, **32**, 211-218.

Chang, C., Rakowski, E., Bearden, J. and Joachimiak, A. (2009). Crystal structure of putative membrane protein from Clostridium difficile 630, Midwest Center for Structural Genomics (MCSG).

Cheknis, A. K., Sambol, S. P., Davidson, D. M., Nagaro, K. J., Mancini, M. C., Hidalgo-Arroyo, G. A., Brazier, J. S., Johnson, S. and Gerding, D. N. (2009). Distribution of Clostridium difficile strains from a North American, European and Australian trial of treatment for C. difficile infections: 2005-2007. *Anaerobe*, **15**, 230-233.

Chen, X., Katchar, K., Goldsmith, J. D., Nanthakumar, N., Cheknis, A., Gerding, D. N. and Kelly, C. P. (2008). A mouse model of Clostridium difficile-associated disease. *Gastroenterology*, **135**, 1984-1992.

Chu, S., Gustafson, C. E., Feutrier, J., Cavaignac, S. and Trust, T. J. (1993). Transcriptional analysis of the Aeromonas salmonicida S-layer protein gene vapA. *J Bacteriol*, **175**, 7968-7975.

Chung, S., Shin, S. H., Bertozzi, C. R. and De Yoreo, J. J. (2010). Self-catalyzed growth of S layers via an amorphous-to-crystalline transition limited by folding kinetics. *Proc Natl Acad Sci U S A*, **107**, 16536-16541.

Clatworthy, A. E., Pierson, E. and Hung, D. T. (2007). Targeting virulence: a new paradigm for antimicrobial therapy. *Nat Chem Biol*, **3**, 541-548.

Claude, J. B., Suhre, K., Notredame, C., Claverie, J. M. and Abergel, C. (2004). CaspR: a web server for automated molecular replacement using homology modelling. *Nucleic Acids Res*, **32**, W606-609.

Combet, C., Jambon, M., Deleage, G. and Geourjon, C. (2002). Geno3D: automatic comparative molecular modelling of protein. *Bioinformatics*, **18**, 213-214.

Corthesy, B. and Spertini, F. (1999). Secretory immunoglobulin A: from mucosal protection to vaccine development. *Biol Chem*, **380**, 1251-1262.

Cousineau, B., Smith, D., Lawrence-Cavanagh, S., Mueller, J. E., Yang, J., Mills, D., Manias, D., Dunny, G., Lambowitz, A. M. and Belfort, M. (1998). Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell*, **94**, 451-462.

Cygler, M. and Mort, J. S. (1997). Proregion structure of members of the papain superfamily. Mode of inhibition of enzymatic activity. *Biochimie*, **79**, 645-652.

Dacanay, A., Boyd, J. M., Fast, M. D., Knickle, L. C. and Reith, M. E. (2010). Aeromonas salmonicida Type I pilus system contributes to host colonization but not invasion. *Dis Aquat Organ*, **88**, 199-206.

Dang, T. H., Riva Lde, L., Fagan, R. P., Storck, E. M., Heal, W. P., Janoir, C., Fairweather, N. F. and Tate, E. W. (2010). Chemical probes of surface layer biogenesis in Clostridium difficile. *ACS Chem Biol*, **5**, 279-285.

Dargatz, H., Diefenthal, T., Witte, V., Reipen, G. and von Wettstein, D. (1993). The heterodimeric protease clostripain from Clostridium histolyticum is encoded by a single gene. *Mol Gen Genet*, **240**, 140-145.

Davies, G. and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure*, **3**, 853-859.

Dawson, L. F., Stabler, R. A. and Wren, B. W. (2008). Assessing the role of p-cresol tolerance in Clostridium difficile. *J Med Microbiol*, **57**, 745-749.

Delmee, M., Avesani, V., Ernest, I. and Surleraux, M. (1990). Detection of specific antigens for ten serogroups of Clostridium difficile. *Mol Cell Probes*, **4**, 1-10.

Delmee, M., Laroche, Y., Avesani, V. and Cornelis, G. (1986). Comparison of serogrouping and polyacrylamide gel electrophoresis for typing Clostridium difficile. *J Clin Microbiol*, **24**, 991-994.

Demarest, S. J., Hariharan, M., Elia, M., Salbato, J., Jin, P., Bird, C., Short, J. M., Kimmel, B. E., Dudley, M., Woodnutt, G. and Hansen, G. (2010). Neutralization of Clostridium difficile toxin A using antibody combinations. *MAbs*, **2**, [Epub ahead of print].

Deneve, C., Bouttier, S., Dupuy, B., Barbut, F., Collignon, A. and Janoir, C. (2009a). Effects of subinhibitory concentrations of antibiotics on colonization factor expression by moxifloxacin-susceptible and moxifloxacin-resistant Clostridium difficile strains. *Antimicrob Agents Chemother*, **53**, 5155-5162.

Deneve, C., Delomenie, C., Barc, M. C., Collignon, A. and Janoir, C. (2008). Antibiotics involved in Clostridium difficile-associated disease increase colonization factor gene expression. *J Med Microbiol*, **57**, 732-738.

Deneve, C., Janoir, C., Poilane, I., Fantinato, C. and Collignon, A. (2009b). New trends in Clostridium difficile virulence and pathogenesis. *Int J Antimicrob Agents*, **33 Suppl 1**, S24-28.

Desvaux, M., Dumas, E., Chafsey, I. and Hebraud, M. (2006). Protein cell surface display in Gram-positive bacteria: from single protein to macromolecular protein structure. *FEMS Microbiol Lett*, **256**, 1-15.

Dineen, S. S., McBride, S. M. and Sonenshein, A. L. (2010). Integration of Metabolism and Virulence by Clostridium difficile CodY. *J Bacteriol*, **192**, 5350-5362.

Dineen, S. S., Villapakkam, A. C., Nordman, J. T. and Sonenshein, A. L. (2007). Repression of Clostridium difficile toxin gene expression by CodY. *Mol Microbiol*, **66**, 206-219.

Doi, R. H. and Kosugi, A. (2004). Cellulosomes: plant-cell-wall-degrading enzyme complexes. *Nat Rev Microbiol*, **2**, 541-551.

Drenth, J., Jansonius, J. N., Koekoek, R., Swen, H. M. and Wolthers, B. G. (1968). Structure of papain. *Nature*, **218**, 929-932.

Drudy, D., Calabi, E., Kyne, L., Sougioultzis, S., Kelly, E., Fairweather, N. and Kelly, C. P. (2004). Human antibody response to surface layer proteins in Clostridium difficile infection. *FEMS Immunol Med Microbiol*, **41**, 237-242.

Dupres, V., Alsteens, D., Pauwels, K. and Dufrene, Y. F. (2009). In vivo imaging of S-layer nanoarrays on Corynebacterium glutamicum. *Langmuir*, **25**, 9653-9655.

Dupuy, B., Govind, R., Antunes, A. and Matamouros, S. (2008). Clostridium difficile toxin synthesis is negatively regulated by TcdC. *J Med Microbiol*, **57**, 685-689.

Edgell, D. R., Belfort, M. and Shub, D. A. (2000). Barriers to intron promiscuity in bacteria. *J Bacteriol*, **182**, 5281-5289.

Eichler, J. (2001). Post-translational modification of the S-layer glycoprotein occurs following translocation across the plasma membrane of the haloarchaeon Haloferax volcanii. *Eur J Biochem*, **268**, 4366-4373.

Eidhin, D. N., Ryan, A. W., Doyle, R. M., Walsh, J. B. and Kelleher, D. (2006). Sequence and phylogenetic analysis of the gene for surface layer protein, slpA, from 14 PCR ribotypes of Clostridium difficile. *J Med Microbiol*, **55**, 69-83.

Elder, B. L., Boraker, D. K. and Fives-Taylor, P. M. (1982). Whole-bacterial cell enzyme-linked immunosorbent assay for Streptococcus sanguis fimbrial antigens. *J Clin Microbiol*, **16**, 141-144.

Elliott, B., Reed, R., Chang, B. J. and Riley, T. V. (2009). Bacteremia with a large clostridial toxin-negative, binary toxin-positive strain of Clostridium difficile. *Anaerobe*, **15**, 249-251.

Emerson, J. E., Reynolds, C. B., Fagan, R. P., Shaw, H. A., Goulding, D. and Fairweather, N. F. (2009a). A novel genetic switch controls phase variable expression of CwpV, a Clostridium difficile cell wall protein. *Mol Microbiol*.

Emerson, J. E., Reynolds, C. B., Fagan, R. P., Shaw, H. A., Goulding, D. and Fairweather, N. F. (2009b). A novel genetic switch controls phase variable expression of CwpV, a Clostridium difficile cell wall protein. *Mol Microbiol*, **74**, 541-556.

Emerson, J. E., Stabler, R. A., Wren, B. W. and Fairweather, N. F. (2008). Microarray analysis of the transcriptional responses of Clostridium difficile to environmental and antibiotic stress. *J Med Microbiol*, **57**, 757-764.

Emsley, P., Lohkamp, B., Scott, W. G. and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, **66**, 486-501.

Engelhardt, H. (2007). Are S-layers exoskeletons? The basic function of protein surface layers revisited. *J Struct Biol*, **160**, 115-124.

Engelhardt, H. and Peters, J. (1998). Structural research on surface layers: a focus on stability, surface layer homology domains, and surface layer-cell wall interactions. *J Struct Biol*, **124**, 276-302.

Enkhtuya, J., Kawamoto, K., Kobayashi, Y., Uchida, I., Rana, N. and Makino, S. (2006). Significant passive protective effect against anthrax by antibody to Bacillus anthracis inactivated spores that lack two virulence plasmids. *Microbiology*, **152**, 3103-3110.

Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U. and Sali, A. (2007). Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci*, **Chapter 2**, Unit 2 9.

Eveillard, M., Fourel, V., Barc, M. C., Kerneis, S., Coconnier, M. H., Karjalainen, T., Bourlioux, P. and Servin, A. L. (1993). Identification and characterization of adhesive factors of Clostridium difficile involved in adhesion to human colonic enterocyte-like Caco-2 and mucus-secreting HT29 cells in culture. *Mol Microbiol*, **7**, 371-381.

Evrard, C. H., Declercq, J. P., Debaerdemaeker, T. and KÖnig, H. (1999). The first successful crystallization of a prokaryotic extremely thermophilic outer surface layer glycoprotein. *Zeitschrift für Kristallographie*, **214**, 427-429.

Ewald, P. (1969). Introduction to the dynamical theory of X-ray diffraction. *Acta Crystallographica Section A*, **25**, 103-108.

Fagan, R. P., Albesa-Jove, D., Qazi, O., Svergun, D. I., Brown, K. A. and Fairweather, N. F. (2009). Structural insights into the molecular organization of the S-layer from Clostridium difficile. *Mol Microbiol*, **71**, 1308-1322.

Fagan, R. P., Janoir, C., Collignon, A., Mastrantonio, P., Poxton, I. R. and Fairweather, N. F. (2011). A proposed nomenclature for cell wall proteins of Clostridium difficile. *J Med Microbiol*, [Epub ahead of print].

Fan, D. P. and Beckman, M. M. (1973). Micrococcus lysodeikticus bacterial walls as a substrate specific for the autolytic glycosidase of Bacillus subtilis. *J Bacteriol*, **114**, 804-813.

Farber, G. K. and Petsko, G. A. (1990). The evolution of alpha/beta barrel enzymes. *Trends Biochem Sci*, **15**, 228-234.

Feese, M. D., Kato, Y., Tamada, T., Kato, M., Komeda, T., Miura, Y., Hirose, M., Hondo, K., Kobayashi, K. and Kuroki, R. (2000). Crystal structure of glycosyltrehalose

trehalohydrolase from the hyperthermophilic archaeum Sulfolobus solfataricus. *J Mol Biol*, **301**, 451-464.

Fein, J. E. and Rogers, H. J. (1976). Autolytic enzyme-deficient mutants of Bacillus subtilis 168. *J Bacteriol*, **127**, 1427-1442.

Felix, C. R. and Ljungdahl, L. G. (1993). The cellulosome: the exocellular organelle of Clostridium. *Annu Rev Microbiol*, **47**, 791-819.

Feltis, B. A., Kim, A. S., Kinneberg, K. M., Lyerly, D. L., Wilkins, T. D., Erlandsen, S. L. and Wells, C. L. (1999). Clostridium difficile toxins may augment bacterial penetration of intestinal epithelium. *Arch Surg*, **134**, 1235-1241; discussion 1241-1232.

Feng, S., Chen, J. K., Yu, H., Simon, J. A. and Schreiber, S. L. (1994). Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. *Science*, **266**, 1241-1247.

Fernandez-Herrero, L. A., Olabarria, G. and Berenguer, J. (1997). Surface proteins and a novel transcription factor regulate the expression of the S-layer gene in Thermus thermophilus HB8. *Mol Microbiol*, **24**, 61-72.

Fernandez, C., Szyperski, T., Bruyere, T., Ramage, P., Mosinger, E. and Wuthrich, K. (1997). NMR solution structure of the pathogenesis-related protein P14a. *J Mol Biol*, **266**, 576-593.

Fernie, D. S., Thomson, R. O., Batty, I. and Walker, P. D. (1983). Active and passive immunization to protect against antibiotic associated caecitis in hamsters. *Dev Biol Stand*, **53**, 325-332.

Frazier, C. L., San Filippo, J., Lambowitz, A. M. and Mills, D. A. (2003). Genetic manipulation of Lactococcus lactis by using targeted group II introns: generation of stable insertions without selection. *Appl Environ Microbiol*, **69**, 1121-1128.

Freeman, J., Baines, S. D., Saxton, K. and Wilcox, M. H. (2007). Effect of metronidazole on growth and toxin production by epidemic Clostridium difficile PCR ribotypes 001 and 027 in a human gut model. *J Antimicrob Chemother*, **60**, 83-91.

Freeman, J., Bauer, M. P., Baines, S. D., Corver, J., Fawley, W. N., Goorhuis, B., Kuijper, E. J. and Wilcox, M. H. (2010). The changing epidemiology of Clostridium difficile infections. *Clin Microbiol Rev*, **23**, 529-549.

Ganeshapillai, J., Vinogradov, E., Rousseau, J., Weese, J. S. and Monteiro, M. A. (2008). Clostridium difficile cell-surface polysaccharides composed of pentaglycosyl and hexaglycosyl phosphate repeating units. *Carbohydr Res*, **343**, 703-710.

Garcia-Nieto, R. M., Rico-Mata, R., Arias-Negrete, S. and Avila, E. E. (2008). Degradation of human secretory IgA1 and IgA2 by Entamoeba histolytica surface-associated proteolytic activity. *Parasitol Int*, **57**, 417-423.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Protocols Handbook*. Walker, J. M., Humana Press**:** 571-607

Gerber, M., Walch, C., Loffler, B., Tischendorf, K., Reischl, U. and Ackermann, G. (2008). Effect of sub-MIC concentrations of metronidazole, vancomycin, clindamycin and linezolid on toxin gene transcription and production in Clostridium difficile. *J Med Microbiol*, **57**, 776-783.

Gerding, D. N. (1997). Is there a relationship between vancomycin-resistant enterococcal infection and Clostridium difficile infection? *Clin Infect Dis*, **25 Suppl 2**, S206-210.

Gerding, D. N. (2010). Global epidemiology of Clostridium difficile infection in 2010. *Infect Control Hosp Epidemiol*, **31 Suppl 1**, S32-34.

Gerding, D. N. and Johnson, S. (2010). Management of Clostridium difficile infection: thinking inside and outside the box. *Clin Infect Dis*, **51**, 1306-1313.

Geric, B., Carman, R. J., Rupnik, M., Genheimer, C. W., Sambol, S. P., Lyerly, D. M., Gerding, D. N. and Johnson, S. (2006). Binary toxin-producing, large clostridial toxin-negative

Clostridium difficile strains are enterotoxic but do not cause disease in hamsters. *J Infect Dis*, **193**, 1143-1150.

Gerstmayr, M., Ilk, N., Schabussova, I., Jahn-Schmid, B., Egelseer, E. M., Sleytr, U. B., Ebner, C. and Bohle, B. (2007). A novel approach to specific allergy treatment: the recombinant allergen-S-layer fusion protein rSbsC-Bet v 1 matures dendritic cells that prime Th0/Th1 and IL-10-producing regulatory T cells. *J Immunol*, **179**, 7270-7275.

Ghantoji, S. S., Sail, K., Lairson, D. R., DuPont, H. L. and Garey, K. W. (2010). Economic healthcare costs of Clostridium difficile infection: a systematic review. *J Hosp Infect*, **74**, 309-318.

Giannasca, P. J., Zhang, Z. X., Lei, W. D., Boden, J. A., Giel, M. A., Monath, T. P. and Thomas, W. D., Jr. (1999). Serum antitoxin antibodies mediate systemic and mucosal protection from Clostridium difficile disease in hamsters. *Infect Immun*, **67**, 527-538.

Giesemann, T., Egerer, M., Jank, T. and Aktories, K. (2008). Processing of Clostridium difficile toxins. *J Med Microbiol*, **57**, 690-696.

Goto, T., Yamashita, A., Hirakawa, H., Matsutani, M., Todo, K., Ohshima, K., Toh, H., Miyamoto, K., Kuhara, S., Hattori, M., Shimizu, T. and Akimoto, S. (2008). Complete genome sequence of Finegoldia magna, an anaerobic opportunistic pathogen. *DNA Res*, **15**, 39-47.

Goulding, D., Thompson, H., Emerson, J., Fairweather, N. F., Dougan, G. and Douce, G. R. (2009). Distinctive profiles of infection and pathology in hamsters infected with Clostridium difficile strains 630 and B1. *Infect Immun*, **77**, 5478-5485.

Grogono-Thomas, R., Blaser, M. J., Ahmadi, M. and Newell, D. G. (2003). Role of S-layer protein antigenic diversity in the immune responses of sheep experimentally challenged with Campylobacter fetus subsp. fetus. *Infect Immun*, **71**, 147-154.

Grogono-Thomas, R., Dworkin, J., Blaser, M. J. and Newell, D. G. (2000). Roles of the surface layer proteins of Campylobacter fetus subsp. fetus in ovine abortion. *Infect Immun*, **68**, 1687-1691.

Guan, S., Bastin, D. A. and Verma, N. K. (1999). Functional analysis of the O antigen glucosylation gene cluster of Shigella flexneri bacteriophage SfX. *Microbiology*, **145 (Pt 5)**, 1263-1273.

Guo, H., Karberg, M., Long, M., Jones, J. P., 3rd, Sullenger, B. and Lambowitz, A. M. (2000). Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. *Science*, **289**, 452-457.

Gyorvary, E. S., Stein, O., Pum, D. and Sleytr, U. B. (2003). Self-assembly and recrystallization of bacterial S-layer proteins at silicon supports imaged in real time by atomic force microscopy. *J Microsc*, **212**, 300-306.

Haiser, H. J., Yousef, M. R. and Elliot, M. A. (2009). Cell wall hydrolases affect germination, vegetative growth, and sporulation in Streptomyces coelicolor. *J Bacteriol*, **191**, 6501-6512.

Hall, I. C. and O'Toole, E. (1935). Intestinal flora in newborn infants with description of a new pathogenic anaerobe. *Am J Dis Child* **49** 390–402.

Hamburger, Z. A., Brown, M. S., Isberg, R. R. and Bjorkman, P. J. (1999). Crystal structure of invasin: a bacterial integrin-binding protein. *Science*, **286**, 291-295.

Hammond, G. A. and Johnson, J. L. (1995). The toxigenic element of Clostridium difficile strain VPI 10463. *Microb Pathog*, **19**, 203-213.

Handfield, M., Brady, L. J., Progulske-Fox, A. and Hillman, J. D. (2000). IVIAT: a novel method to identify microbial genes expressed specifically during human infections. *Trends Microbiol*, **8**, 336-339.

Hautefort, I., Hinton, J. C. D. and Philippe Sansonetti, A. Z. (2002). 4 Molecular methods for monitoring bacterial gene expression during infection. *Methods in Microbiology*, Academic Press. **Volume 31:** 55-91.

He, M., Sebaihia, M., Lawley, T. D., Stabler, R. A., Dawson, L. F., Martin, M. J., Holt, K. E., Seth-Smith, H. M., Quail, M. A., Rance, R., Brooks, K., Churcher, C., Harris, D., Bentley, S. D., Burrows, C., Clark, L., Corton, C., Murray, V., Rose, G., Thurston, S., van Tonder, A., Walker, D., Wren, B. W., Dougan, G. and Parkhill, J. (2010). Evolutionary dynamics of Clostridium difficile over short and long time scales. *Proc Natl Acad Sci U S A*, **107**, 7527-7532.

Heap, J. T., Cartman, S. T., Pennington, O. J., Cooksley, C. M., Scott, J. C., Blount, B., Burns, D. and Minton, N. P. (2009a). Development of Genetic Knock-out Systems for Clostridia *Clostridia: Molecular Biology in the Post-genomic Era*. Brüggemann, H. and Gottschalk, G., Caister Academic Press.

Heap, J. T., Kuehne, S. A., Ehsaan, M., Cartman, S. T., Cooksley, C. M., Scott, J. C. and Minton, N. P. (2010). The ClosTron: Mutagenesis in Clostridium refined and streamlined. *J Microbiol Methods*, **80**, 49-55.

Heap, J. T., Pennington, O. J., Cartman, S. T., Carter, G. P. and Minton, N. P. (2007). The ClosTron: a universal gene knock-out system for the genus Clostridium. *J Microbiol Methods*, **70**, 452-464.

Heap, J. T., Pennington, O. J., Cartman, S. T. and Minton, N. P. (2009b). A modular system for Clostridium shuttle plasmids. *J Microbiol Methods*, **78**, 79-85.

Hendrickson, W. A., Ogata, C. M. and Charles W. Carter, Jr. (1997). [28] Phase determination from multiwavelength anomalous diffraction measurements. *Methods in Enzymology*, Academic Press. **Volume 276:** 494-523.

Hennequin, C., Janoir, C., Barc, M. C., Collignon, A. and Karjalainen, T. (2003). Identification and characterization of a fibronectin-binding protein from Clostridium difficile. *Microbiology*, **149**, 2779-2787.

Hennequin, C., Porcheray, F., Waligora-Dupriet, A., Collignon, A., Barc, M., Bourlioux, P. and Karjalainen, T. (2001). GroEL (Hsp60) of Clostridium difficile is involved in cell adherence. *Microbiology*, **147**, 87-96.

Henrissat, B. and Davies, G. (1997). Structural and sequence-based classification of glycoside hydrolases. *Curr Opin Struct Biol*, **7**, 637-644.

Hensel, M., Shea, J. E., Gleeson, C., Jones, M. D., Dalton, E. and Holden, D. W. (1995). Simultaneous identification of bacterial virulence genes by negative selection. *Science*, **269**, 400-403.

Hermann, J. C., Marti-Arbona, R., Fedorov, A. A., Fedorov, E., Almo, S. C., Shoichet, B. K. and Raushel, F. M. (2007). Structure-based activity prediction for an enzyme of unknown function. *Nature*, **448**, 775-779.

Ho, J. G., Greco, A., Rupnik, M. and Ng, K. K. (2005). Crystal structure of receptor-binding C-terminal repeats from Clostridium difficile toxin A. *Proc Natl Acad Sci U S A*, **102**, 18373-18378.

Holmgren, J. and Czerkinsky, C. (2005). Mucosal immunity and vaccines. *Nat Med*, **11**, S45-53.

Holmgren, J., Czerkinsky, C., Lycke, N. and Svennerholm, A. M. (1992). Mucosal immunity: implications for vaccine development. *Immunobiology*, **184**, 157-179.

Horvath, M. P. and Schultz, S. C. (2001). DNA G-quartets in a 1.86 A resolution structure of an Oxytricha nova telomeric protein-DNA complex. *J Mol Biol*, **310**, 367-377.

Howerton, A., Ramirez, N. and Abel-Santos, E. (2011). Mapping Interactions between Germinants and Clostridium difficile Spores. *J Bacteriol*, **193**, 274-282.

HPA (2008). Processing of faeces for Clostridium difficile BSOP 10. http://www.hpa-standardmethods.org.uk/, Health Protection Agency.

Huang, W., Matte, A., Li, Y., Kim, Y. S., Linhardt, R. J., Su, H. and Cygler, M. (1999). Crystal structure of chondroitinase B from Flavobacterium heparinum and its complex with a disaccharide product at 1.7 A resolution. *J Mol Biol*, **294**, 1257-1269.

Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L., Finn, R. D., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Laugraud, A., Letunic, I., Lonsdale, D., Lopez, R., Madera, M., Maslen, J., McAnulla, C., McDowall, J., Mistry, J., Mitchell, A., Mulder, N., Natale, D., Orengo, C., Quinn, A. F., Selengut, J. D., Sigrist, C. J., Thimma, M., Thomas, P. D., Valentin, F., Wilson, D., Wu, C. H. and Yeats, C. (2009). InterPro: the integrative protein signature database. *Nucleic Acids Res*, **37**, D211-215.

Hussain, H. A., Roberts, A. P. and Mullany, P. (2005). Generation of an erythromycin-sensitive derivative of Clostridium difficile strain 630 (630Deltaerm) and demonstration that the conjugative transposon Tn916DeltaE enters the genome of this strain at multiple sites. *J Med Microbiol*, **54**, 137-141.

Hussain, H. A., Roberts, A. P., Whalan, R. and Mullany, P. (2010). Transposon mutagenesis in Clostridium difficile. *Methods Mol Biol*, **646**, 203-211.

Hutchings, M. I., Palmer, T., Harrington, D. J. and Sutcliffe, I. C. (2009). Lipoprotein biogenesis in Gram-positive bacteria: knowing when to hold 'em, knowing when to fold 'em. *Trends Microbiol*, **17**, 13-21.

Hytonen, J., Haataja, S. and Finne, J. (2003). Streptococcus pyogenes glycoprotein-binding strepadhesin activity is mediated by a surface-associated carbohydrate-degrading enzyme, pullulanase. *Infect Immun*, **71**, 784-793.

Ireton, K. and Cossart, P. (1998). Interaction of invasive bacteria with host signaling pathways. *Curr Opin Cell Biol*, **10**, 276-283.

Ishiguro, E. E., Kay, W. W., Ainsworth, T., Chamberlain, J. B., Austen, R. A., Buckley, J. T. and Trust, T. J. (1981). Loss of virulence during culture of Aeromonas salmonicida at high temperature. *J Bacteriol*, **148**, 333-340.

Janda, J. M., Kokka, R. P. and Guthertz, L. S. (1994). The susceptibility of S-layer-positive and S-layer-negative Aeromonas strains to complement-mediated lysis. *Microbiology*, **140 (Pt 10)**, 2899-2905.

Janecek, S., Svensson, B. and MacGregor, E. A. (2007). A remote but significant sequence homology between glycoside hydrolase clan GH-H and family GH31. *FEBS Lett*, **581**, 1261-1268.

Janoir, C., Grenery, J., Savariau-Lacomme, M. P. and Collignon, A. (2004). [Characterization of an extracellular protease from Clostridium difficile]. *Pathol Biol (Paris)*, **52**, 444-449.

Janoir, C., Pechine, S., Grosdidier, C. and Collignon, A. (2007). Cwp84, a surface-associated protein of Clostridium difficile is a cysteine protease with degrading activity on extracellular matrix proteins. *J Bacteriol*, **189**, 7174-7180.

Jarchum, I., Liu, M., Lipuma, L. and Pamer, E. G. (2011). Toll-like receptor-5 stimulation protects mice from acute Clostridium difficile colitis. *Infect Immun*, **79**, 1498-1503.

Jarrell, K. F. and McBride, M. J. (2008). The surprisingly diverse ways that prokaryotes move. *Nat Rev Microbiol*, **6**, 466-476.

Jenkins, W. T. (1998). Three solutions of the protein solubility problem. *Protein Sci*, **7**, 376-382.

Jia, K., Zhu, Y., Zhang, Y. and Li, Y. (2011). Group II Intron-Anchored Gene Deletion in Clostridium. *PLoS One*, **6**, e16693.

Johnson, S. (1997). Antibody responses to clostridial infection in humans. *Clin Infect Dis*, **25 Suppl 2**, S173-177.

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol*, **292**, 195-202.

Jonquieres, R., Bierne, H., Fiedler, F., Gounon, P. and Cossart, P. (1999). Interaction between the protein InlB of Listeria monocytogenes and lipoteichoic acid: a novel mechanism of protein association at the surface of gram-positive bacteria. *Mol Microbiol*, **34**, 902-914.

Jonquières, R., Bierne, H., Fiedler, F., Gounon, P. and Cossart, P. (1999). Interaction between the protein InlB of Listeria monocytogenes and lipoteichoic acid: a novel mechanism of protein association at the surface of Gram-positive bacteria. *Molecular Microbiology*, **34**, 902-914.

Juang, P., Skledar, S. J., Zgheib, N. K., Paterson, D. L., Vergis, E. N., Shannon, W. D., Ansani, N. T. and Branch, R. A. (2007). Clinical outcomes of intravenous immune globulin in severe clostridium difficile-associated diarrhea. *Am J Infect Control*, **35**, 131-137.

Kahala, M., Savijoki, K. and Palva, A. (1997). In vivo expression of the Lactobacillus brevis S-layer gene. *J Bacteriol*, **179**, 284-286.

Karberg, M., Guo, H., Zhong, J., Coon, R., Perutka, J. and Lambowitz, A. M. (2001). Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. *Nat Biotechnol*, **19**, 1162-1167.

Karjalainen, T., Barc, M. C., Collignon, A., Trolle, S., Boureau, H., Cotte-Laffitte, J. and Bourlioux, P. (1994). Cloning of a genetic determinant from Clostridium difficile involved in adherence to tissue culture cells and mucus. *Infect Immun*, **62**, 4347-4355.

Karjalainen, T., Saumier, N., Barc, M. C., Delmee, M. and Collignon, A. (2002). Clostridium difficile genotyping based on slpA variable region in S-layer gene sequence: an alternative to serotyping. *J Clin Microbiol*, **40**, 2452-2458.

Karjalainen, T., Waligora-Dupriet, A. J., Cerquetti, M., Spigaglia, P., Maggioni, A., Mauri, P. and Mastrantonio, P. (2001). Molecular and genomic analysis of genes encoding surface-anchored proteins from Clostridium difficile. *Infect Immun*, **69**, 3442-3446.

Karlsson, S., Burman, L. G. and Akerlund, T. (2008). Induction of toxins in Clostridium difficile is associated with dramatic changes of its metabolism. *Microbiology*, **154**, 3430-3436.

Karlsson, S., Dupuy, B., Mukherjee, K., Norin, E., Burman, L. G. and Akerlund, T. (2003). Expression of Clostridium difficile toxins A and B and their sigma factor TcdD is controlled by temperature. *Infect Immun*, **71**, 1784-1793.

Katchar, K., Taylor, C. P., Tummala, S., Chen, X., Sheikh, J. and Kelly, C. P. (2007). Association between IgG2 and IgG3 subclass responses to toxin A and recurrent Clostridium difficile-associated disease. *Clin Gastroenterol Hepatol*, **5**, 707-713.

Kato, H., Kato, H., Nakamura, M., Iwashima, Y., Nakamura, A. and Ueda, R. (2009). Rapid analysis of Clostridium difficile strains recovered from hospitalized patients by using the slpA sequence typing system. *J Infect Chemother*, **15**, 199-202.

Kato, H., Yokoyama, T. and Arakawa, Y. (2005). Typing by sequencing the slpA gene of Clostridium difficile strains causing multiple outbreaks in Japan. *J Med Microbiol*, **54**, 167-171.

Kawata, T., Takeoka, A., Takumi, K. and Masuda, K. (1984). Demonstration and preliminary characterization of a regular array in the cell wall of Clostridium difficile. *FEMS Microbiology Letters*, **24**, 323-328.

Kay, W. W. and Trust, T. J. (1991). Form and functions of the regular surface array (S-layer) of Aeromonas salmonicida. *Experientia*, **47**, 412-414.

Keel, M. K. and Songer, J. G. (2006). The comparative pathology of Clostridium difficile-associated disease. *Vet Pathol*, **43**, 225-240.

Kelley, L. A. and Sternberg, M. J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc*, **4**, 363-371.

Kelly, C. P. and LaMont, J. T. (1998). Clostridium difficile infection. *Annu Rev Med*, **49**, 375-390.

Kelly, G., Prasannan, S., Daniell, S., Fleming, K., Frankel, G., Dougan, G., Connerton, I. and Matthews, S. (1999). Structure of the cell-adhesion fragment of intimin from enteropathogenic Escherichia coli. *Nat Struct Biol*, **6**, 313-318.

Kerff, F., Petrella, S., Mercier, F., Sauvage, E., Herman, R., Pennartz, A., Zervosen, A., Luxen, A., Frere, J. M., Joris, B. and Charlier, P. (2010). Specific structural features of the N-acetylmuramoyl-L-alanine amidase AmiD from Escherichia coli and mechanistic implications for enzymes of this family. *J Mol Biol*, **397**, 249-259.

Kern, J., Ryan, C., Faull, K. and Schneewind, O. (2010). Bacillus anthracis surface-layer proteins assemble by binding to the secondary cell wall polysaccharide in a manner that requires csaB and tagO. *J Mol Biol*, **401**, 757-775.

Kern, J. and Schneewind, O. (2009). BslA, the S-layer adhesin of B. anthracis, is a virulence factor for anthrax pathogenesis. *Mol Microbiol*.

Killgore, G., Thompson, A., Johnson, S., Brazier, J., Kuijper, E., Pepin, J., Frost, E. H., Savelkoul, P., Nicholson, B., van den Berg, R. J., Kato, H., Sambol, S. P., Zukowski, W., Woods, C., Limbago, B., Gerding, D. N. and McDonald, L. C. (2008). Comparison of seven techniques for typing international epidemic strains of Clostridium difficile: restriction endonuclease analysis, pulsed-field gel electrophoresis, PCR-ribotyping, multilocus sequence typing, multilocus variable-number tandem-repeat analysis, amplified fragment length polymorphism, and surface layer protein A gene sequence typing. *J Clin Microbiol*, **46**, 431-437.

Kim, K., Pickering, L. K., DuPont, H. L., Sullivan, N. and Wilkins, T. (1984). In vitro and in vivo neutralizing activity of human colostrum and milk against purified toxins A and B of Clostridium difficile. *J Infect Dis*, **150**, 57-62.

Kim, K. H., Fekety, R., Batts, D. H., Brown, D., Cudmore, M., Silva, J., Jr. and Waters, D. (1981). Isolation of Clostridium difficile from the environment and contacts of patients with antibiotic-associated colitis. *J Infect Dis*, **143**, 42-50.

Kim, M., Ashida, H., Ogawa, M., Yoshikawa, Y., Mimuro, H. and Sasakawa, C. (2010). Bacterial interactions with the host epithelium. *Cell Host Microbe*, **8**, 20-35.

Kink, J. A. and Williams, J. A. (1998). Antibodies to recombinant Clostridium difficile toxins A and B are an effective treatment and prevent relapse of C. difficile-associated disease in a hamster model of infection. *Infect Immun*, **66**, 2018-2025.

Kinoshita, T. and Inoue, N. (2000). Dissecting and manipulating the pathway for glycosylphos-phatidylinositol-anchor biosynthesis. *Curr Opin Chem Biol*, **4**, 632-638.

Kirby, J. M., Ahern, H., Roberts, A. K., Kumar, V., Freeman, Z., Acharya, K. R. and Shone, C. C. (2009). Cwp84, a surface-associated cysteine protease, plays a role in the maturation of the surface layer of Clostridium difficile. *J Biol Chem*, **284**, 34666-34673.

Kline, K. A., Dodson, K. W., Caparon, M. G. and Hultgren, S. J. (2010). A tale of two pili: assembly and function of pili in bacteria. *Trends Microbiol*, **18**, 224-232.

Kodama, T., Endo, K., Ara, K., Ozaki, K., Kakeshita, H., Yamane, K. and Sekiguchi, J. (2007). Effect of Bacillus subtilis spo0A mutation on cell wall lytic enzymes and extracellular proteases, and prevention of cell lysis. *J Biosci Bioeng*, **103**, 13-21.

Konstantinov, S. R., Smidt, H., de Vos, W. M., Bruijns, S. C., Singh, S. K., Valence, F., Molle, D., Lortal, S., Altermann, E., Klaenhammer, T. R. and van Kooyk, Y. (2008). S layer protein A of Lactobacillus acidophilus NCFM regulates immature dendritic cell and T cell functions. *Proc Natl Acad Sci U S A*, **105**, 19474-19479.

Koo, H. L., Garey, K. W. and Dupont, H. L. (2010). Future novel therapeutic agents for Clostridium difficile infection. *Expert Opin Investig Drugs*, **19**, 825-836.

Koshland, D. E. (1953). STEREOCHEMISTRY AND THE MECHANISM OF ENZYMATIC REACTIONS. *Biological Reviews*, **28**, 416-436.

Kotloff, K. L., Wasserman, S. S., Losonsky, G. A., Thomas, W., Jr., Nichols, R., Edelman, R., Bridwell, M. and Monath, T. P. (2001). Safety and immunogenicity of increasing doses of a Clostridium difficile toxoid vaccine administered to healthy adults. *Infect Immun*, **69**, 988-995.

Koulaouzidis, A., Tatham, R., Moschos, J. and Tan, C. W. (2008). Successful treatment of Clostridium difficile colitis with intravenous immunoglobulin. *J Gastrointestin Liver Dis*, **17**, 353-355.

Krupa, J. C. and Mort, J. S. (2000). Optimization of detergents for the assay of cathepsins B, L, S, and K. *Anal Biochem*, **283**, 99-103.

Kuehne, S. A., Cartman, S. T., Heap, J. T., Kelly, M. L., Cockayne, A. and Minton, N. P. (2010). The role of toxin A and toxin B in Clostridium difficile infection. *Nature*.

Kuroda, A. and Sekiguchi, J. (1991). Molecular cloning and sequencing of a major Bacillus subtilis autolysin gene. *J Bacteriol*, **173**, 7304-7312.

Kyne, L., Warny, M., Qamar, A. and Kelly, C. P. (2000). Asymptomatic carriage of Clostridium difficile and serum levels of IgG antibody against toxin A. *N Engl J Med*, **342**, 390-397.

Kyne, L., Warny, M., Qamar, A. and Kelly, C. P. (2001). Association between antibody response to toxin A and protection against recurrent Clostridium difficile diarrhoea. *Lancet*, **357**, 189-193.

Lambert, C., Leonard, N., De Bolle, X. and Depiereux, E. (2002). ESyPred3D: Prediction of proteins 3D structures. *Bioinformatics*, **18**, 1250-1256.

Lambowitz, A. M. and Zimmerly, S. (2004). Mobile group II introns. *Annu Rev Genet*, **38**, 1-35.

Landon, J., Chard, T. and Coxon, R. (1995). Therapeutic antibodies, Springer.

Larson, H. E., Barclay, F. E., Honour, P. and Hill, I. D. (1982). Epidemiology of Clostridium difficile in infants. *J Infect Dis*, **146**, 727-733.

Lasa, I., Caston, J. R., Fernandez-Herrero, L. A., de Pedro, M. A. and Berenguer, J. (1992). Insertional mutagenesis in the extreme thermophilic eubacteria Thermus thermophilus HB8. *Mol Microbiol*, **6**, 1555-1564.

Lavonas, E. J., Schaeffer, T. H., Kokko, J., Mlynarchek, S. L. and Bogdan, G. M. (2009). Crotaline Fab antivenom appears to be effective in cases of severe North American pit viper envenomation: an integrative review. *BMC Emerg Med*, **9**, 13.

Lawley, T. D., Clare, S., Walker, A. W., Goulding, D., Stabler, R. A., Croucher, N., Mastroeni, P., Scott, P., Raisen, C., Mottram, L., Fairweather, N. F., Wren, B. W., Parkhill, J. and Dougan, G. (2009a). Antibiotic treatment of clostridium difficile carrier mice triggers a supershedder state, spore-mediated transmission, and severe disease in immunocompromised hosts. *Infect Immun*, **77**, 3661-3669.

Lawley, T. D., Croucher, N. J., Yu, L., Clare, S., Sebaihia, M., Goulding, D., Pickard, D. J., Parkhill, J., Choudhary, J. and Dougan, G. (2009b). Proteomic and genomic characterization of highly infectious Clostridium difficile 630 spores. *J Bacteriol*, **191**, 5377-5386.

Lazarevic, V., Margot, P., Soldo, B. and Karamata, D. (1992). Sequencing and analysis of the Bacillus subtilis lytRABC divergon: a regulatory unit encompassing the structural genes of the N-acetylmuramoyl-L-alanine amidase and its modifier. *J Gen Microbiol*, **138**, 1949-1961.

Leav, B. A., Blair, B., Leney, M., Knauber, M., Reilly, C., Lowy, I., Gerding, D. N., Kelly, C. P., Katchar, K., Baxter, R., Ambrosino, D. and Molrine, D. (2010). Serum anti-toxin B antibody correlates with protection from recurrent Clostridium difficile infection (CDI). *Vaccine*, **28**, 965-969.

Lee, B. Y., Popovich, M. J., Tian, Y., Bailey, R. R., Ufberg, P. J., Wiringa, A. E. and Muder, R. R. (2010a). The potential value of Clostridium difficile vaccine: an economic computer simulation model. *Vaccine*, **28**, 5245-5253.

Lee, H. S., Kim, M. S., Cho, H. S., Kim, J. I., Kim, T. J., Choi, J. H., Park, C., Lee, H. S., Oh, B. H. and Park, K. H. (2002a). Cyclomaltodextrinase, neopullulanase, and maltogenic amylase are nearly indistinguishable from each other. *J Biol Chem*, **277**, 21891-21897.

Lee, K. Y., Hyeok Yoon, J. H., Kim, M., Roh, S., Lee, Y. S., Seong, B. L. and Kim, K. (2002b). A dipalmitoyl peptide that binds SH3 domain, disturbs intracellular signal transduction, and inhibits tumor growth in vivo. *Biochem Biophys Res Commun*, **296**, 434-442.

Lee, S. M., Lee, J. Y., Park, K. J., Park, J. S., Ha, U. H., Kim, Y. and Lee, H. S. (2010b). The regulator RamA influences cmytA transcription and cell morphology of *Corynebacterium ammoniagenes*. *Curr Microbiol*, **61**, 92-100.

Lemee, L., Bourgeois, I., Ruffin, E., Collignon, A., Lemeland, J. F. and Pons, J. L. (2005). Multilocus sequence analysis and comparative evolution of virulence-associated genes and housekeeping genes of Clostridium difficile. *Microbiology*, **151**, 3171-3180.

Leslie, A. G. (2006). The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr*, **62**, 48-57.

Leung, D. Y., Kelly, C. P., Boguniewicz, M., Pothoulakis, C., LaMont, J. T. and Flores, A. (1991). Treatment with intravenously administered gamma globulin of chronic relapsing colitis induced by Clostridium difficile toxin. *J Pediatr*, **118**, 633-637.

Li, S. S. (2005). Specificity and versatility of SH3 and other proline-recognition domains: structural basis and implications for cellular signal transduction. *Biochem J*, **390**, 641-653.

Lin, Y. P., Kuo, C. J., Koleci, X., McDonough, S. P. and Chang, Y. F. (2010). Manganese binds to Clostrdium difficile Fbp68 and is essential for fibronectin binding. *J Biol Chem*, **286**, 3957-3969.

Lin, Y. P., Raman, R., Sharma, Y. and Chang, Y. F. (2008). Calcium binds to leptospiral immunoglobulin-like protein, LigB, and modulates fibronectin binding. *J Biol Chem*, **283**, 25140-25149.

Linevsky, J. K., Pothoulakis, C., Keates, S., Warny, M., Keates, A. C., Lamont, J. T. and Kelly, C. P. (1997). IL-8 release and neutrophil activation by Clostridium difficile toxin-exposed human monocytes. *Am J Physiol*, **273**, G1333-1340.

Liu, C., Mao, K., Zhang, M., Sun, Z., Hong, W., Li, C., Peng, B. and Chang, Z. (2008a). The SH3-like domain switches its interaction partners to modulate the repression activity of mycobacterial iron-dependent transcription regulator in response to metal ion fluctuations. *J Biol Chem*, **283**, 2439-2453.

Liu, M., Li, S., Hu, S., Zhao, C., Bi, D. and Sun, M. (2008b). Display of avian influenza virus nucleoprotein on Bacillus thuringiensis cell surface using CTC as a fusion partner. *Appl Microbiol Biotechnol*, **78**, 669-676.

Liyanage, H., Kashket, S., Young, M. and Kashket, E. R. (2001). Clostridium beijerinckii and Clostridium difficile detoxify methylglyoxal by a novel mechanism involving glycerol dehydrogenase. *Appl Environ Microbiol*, **67**, 2004-2010.

Lobley, A., Sadowski, M. I. and Jones, D. T. (2009). pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination. *Bioinformatics*, **25**, 1761-1767.

Lowy, I., Molrine, D. C., Leav, B. A., Blair, B. M., Baxter, R., Gerding, D. N., Nichol, G., Thomas, W. D., Jr., Leney, M., Sloan, S., Hay, C. A. and Ambrosino, D. M. (2010).

Treatment with monoclonal antibodies against Clostridium difficile toxins. *N Engl J Med*, **362**, 197-205.

Lu, X. L., Cao, X., Liu, X. Y. and Jiao, B. H. (2010). Recent progress of Src SH2 and SH3 inhibitors as anticancer agents. *Curr Med Chem*, **17**, 1117-1124.

Luo, Y., Frey, E. A., Pfuetzner, R. A., Creagh, A. L., Knoechel, D. G., Haynes, C. A., Finlay, B. B. and Strynadka, N. C. (2000). Crystal structure of enteropathogenic Escherichia coli intimin-receptor complex. *Nature*, **405**, 1073-1077.

Lupas, A. (1996). A circular permutation event in the evolution of the SLH domain? *Mol Microbiol*, **20**, 897-898.

Lupas, A., Engelhardt, H., Peters, J., Santarius, U., Volker, S. and Baumeister, W. (1994). Domain structure of the Acetogenium kivui surface layer revealed by electron crystallography and sequence analysis. *J Bacteriol*, **176**, 1224-1233.

Lyerly, D. M., Bostwick, E. F., Binion, S. B. and Wilkins, T. D. (1991). Passive immunization of hamsters against disease caused by Clostridium difficile by use of bovine immunoglobulin G concentrate. *Infect Immun*, **59**, 2215-2218.

Lyras, D., O'Connor, J. R., Howarth, P. M., Sambol, S. P., Carter, G. P., Phumoonna, T., Poon, R., Adams, V., Vedantam, G., Johnson, S., Gerding, D. N. and Rood, J. I. (2009). Toxin B is essential for virulence of Clostridium difficile. *Nature*, **458**, 1176-1179.

MacCannell, D. R., Louie, T. J., Gregson, D. B., Laverdiere, M., Labbe, A. C., Laing, F. and Henwick, S. (2006). Molecular analysis of Clostridium difficile PCR ribotype 027 isolates from Eastern and Western Canada. *J Clin Microbiol*, **44**, 2147-2152.

MacGregor, E. A., Janecek, S. and Svensson, B. (2001). Relationship of sequence and structure to specificity in the alpha-amylase family of enzymes. *Biochim Biophys Acta*, **1546**, 1-20.

Macnab, R. M. (2003). How bacteria assemble flagella. *Annu Rev Microbiol*, **57**, 77-100.

Maeda, H. (1996). Role of microbial proteases in pathogenesis. *Microbiol Immunol*, **40**, 685-699.

Mallozzi, M., Bozue, J., Giorno, R., Moody, K. S., Slack, A., Cote, C., Qiu, D., Wang, R., McKenney, P., Lai, E. M., Maddock, J. R., Friedlander, A., Welkos, S., Eichenberger, P. and Driks, A. (2008). Characterization of a Bacillus anthracis spore coat-surface protein that influences coat-surface morphology. *FEMS Microbiol Lett*, **289**, 110-117.

Mani, N., Baddour, L. M., Offutt, D. Q., Vijaranakul, U., Nadakavukaren, M. J. and Jayaswal, R. K. (1994). Autolysis-defective mutant of Staphylococcus aureus: pathological considerations, genetic mapping, and electron microscopic studies. *Infect Immun*, **62**, 1406-1409.

Marblestone, J. G., Edavettal, S. C., Lim, Y., Lim, P., Zuo, X. and Butt, T. R. (2006). Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein Sci*, **15**, 182-189.

Marches, O., Batchelor, M., Shaw, R. K., Patel, A., Cummings, N., Nagai, T., Sasakawa, C., Carlsson, S. R., Lundmark, R., Cougoule, C., Caron, E., Knutton, S., Connerton, I. and Frankel, G. (2006). EspF of enteropathogenic Escherichia coli binds sorting nexin 9. *J Bacteriol*, **188**, 3110-3115.

Marchler-Bauer, A., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., Geer, L. Y., Geer, R. C., Gonzales, N. R., Gwadz, M., He, S., Hurwitz, D. I., Jackson, J. D., Ke, Z., Lanczycki, C. J., Liebert, C. A., Liu, C., Lu, F., Lu, S., Marchler, G. H., Mullokandov, M., Song, J. S., Tasneem, A., Thanki, N., Yamashita, R. A., Zhang, D., Zhang, N. and Bryant, S. H. (2009). CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res*, **37**, D205-210.

Marino, M., Banerjee, M., Jonquieres, R., Cossart, P. and Ghosh, P. (2002). GW domains of the Listeria monocytogenes invasion protein InlB are SH3-like and mediate binding to host ligands. *Embo J*, **21**, 5623-5634.

Marion, C., Limoli, D. H., Bobulsky, G. S., Abraham, J. L., Burnaugh, A. M. and King, S. J. (2009). Identification of a pneumococcal glycosidase that modifies O-linked glycans. *Infect Immun*, **77**, 1389-1396.

Marraffini, L. A., Dedent, A. C. and Schneewind, O. (2006). Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. *Microbiol Mol Biol Rev*, **70**, 192-221.

Matsuura, M., Saldanha, R., Ma, H., Wank, H., Yang, J., Mohr, G., Cavanagh, S., Dunny, G. M., Belfort, M. and Lambowitz, A. M. (1997). A bacterial group II intron encoding reverse transcriptase, maturase, and DNA endonuclease activities: biochemical demonstration of maturase activity and insertion of new genetic information within the intron. *Genes Dev*, **11**, 2910-2924.

Matsuura, Y. (2002). A possible mechanism of catalysis involving three essential residues in the enzymes of alpha-amylase family. *Biologia* **57 (Suppl. 11)** 21–27.

Matthews, B. W. (1968). Solvent content of protein crystals. *J Mol Biol*, **33**, 491-497.

Maubach, G., Schilling, K., Rommerskirch, W., Wenz, I., Schultz, J. E., Weber, E. and Wiederanders, B. (1997). The inhibition of cathepsin S by its propeptide--specificity and mechanism of action. *Eur J Biochem*, **250**, 745-750.

Mauri, P. L., Pietta, P. G., Maggioni, A., Cerquetti, M., Sebastianelli, A. and Mastrantonio, P. (1999). Characterization of surface layer proteins from Clostridium difficile by liquid chromatography/electrospray ionization mass spectrometry. *Rapid Commun Mass Spectrom*, **13**, 695-703.

Mayer, B. J. (2001). SH3 domains: complexity in moderation. *J Cell Sci*, **114**, 1253-1263.

McCoubrey, J. and Poxton, I. R. (2001). Variation in the surface layer proteins of Clostridium difficile. *FEMS Immunol Med Microbiol*, **31**, 131-135.

McCoy, A. J. (2005). "Phase Diagrams." from http://www-structmed.cimr.cam.ac.uk/Course/Crystals/Theory/phases.html.

McFarland, L. V. (2002). What's lurking under the bed? Persistence and predominance of particular Clostridium difficile strains in a hospital and the potential role of environmental contamination. *Infect Control Hosp Epidemiol*, **23**, 639-640.

McFarland, L. V. (2005). Alternative treatments for Clostridium difficile disease: what really works? *J Med Microbiol*, **54**, 101-111.

McFarland, L. V. (2008). Update on the changing epidemiology of Clostridium difficile-associated disease. *Nat Clin Pract Gastroenterol Hepatol*, **5**, 40-48.

Mercier, C., Durrieu, C., Briandet, R., Domakova, E., Tremblay, J., Buist, G. and Kulakauskas, S. (2002). Positive role of peptidoglycan breaks in lactococcal biofilm formation. *Mol Microbiol*, **46**, 235-243.

Merrigan, M., Venugopal, A., Mallozzi, M., Roxas, B., Viswanathan, V. K., Johnson, S., Gerding, D. N. and Vedantam, G. (2010). Human hypervirulent Clostridium difficile strains exhibit increased sporulation as well as robust toxin production. *J Bacteriol*, **192**, 4904-4911.

Mesnage, S., Fontaine, T., Mignot, T., Delepierre, M., Mock, M. and Fouet, A. (2000). Bacterial SLH domain proteins are non-covalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation. *Embo J*, **19**, 4473-4484.

Mesnage, S., Tosi-Couture, E., Gounon, P., Mock, M. and Fouet, A. (1998). The capsule and S-layer: two independent and yet compatible macromolecular structures in Bacillus anthracis. *J Bacteriol*, **180**, 52-58.

Messner, P. and Sleytr, U. B. (1992). Crystalline bacterial cell-surface layers. *Adv Microb Physiol*, **33**, 213-275.

Miller, W. G., Bates, A. H., Horn, S. T., Brandl, M. T., Wachtel, M. R. and Mandrell, R. E. (2000). Detection on surfaces and in Caco-2 cells of Campylobacter jejuni cells transformed with new gfp, yfp, and cfp marker plasmids. *Appl Environ Microbiol*, **66**, 5426-5436.

Mukherjee, K., Karlsson, S., Burman, L. G. and Akerlund, T. (2002). Proteins released during high toxin production in Clostridium difficile. *Microbiology*, **148**, 2245-2253.

Muller, D. J., Baumeister, W. and Engel, A. (1996). Conformational change of the hexagonally packed intermediate layer of Deinococcus radiodurans monitored by atomic force microscopy. *J Bacteriol*, **178**, 3025-3030.

Munn, C. B., Ishiguro, E. E., Kay, W. W. and Trust, T. J. (1982). Role of surface components in serum resistance of virulent Aeromonas salmonicida. *Infect Immun*, **36**, 1069-1075.

Murzin, A. G., Brenner, S. E., Hubbard, T. and Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol*, **247**, 536-540.

Nagano, N., Orengo, C. A. and Thornton, J. M. (2002). One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J Mol Biol*, **321**, 741-765.

Neutra, M. R. and Kozlowski, P. A. (2006). Mucosal vaccines: the promise and the challenge. *Nat Rev Immunol*, **6**, 148-158.

Ng, J., Hirota, S. A., Gross, O., Li, Y., Ulke-Lemee, A., Potentier, M. S., Schenck, L. P., Vilaysane, A., Seamone, M. E., Feng, H., Armstrong, G. D., Tschopp, J., Macdonald, J. A., Muruve, D. A. and Beck, P. L. (2010). Clostridium difficile toxin-induced inflammation and intestinal injury are mediated by the inflammasome. *Gastroenterology*, **139**, 542-552, 552 e541-543.

Nguyen, J. T., Turck, C. W., Cohen, F. E., Zuckermann, R. N. and Lim, W. A. (1998). Exploiting the basis of proline recognition by SH3 and WW domains: design of N-substituted inhibitors. *Science*, **282**, 2088-2092.

Ni Eidhin, D. B., O'Brien, J. B., McCabe, M. S., Athie-Morales, V. and Kelleher, D. P. (2008). Active immunization of hamsters against Clostridium difficile infection using surface-layer protein. *FEMS Immunol Med Microbiol*, **52**, 207-218.

Nielsen, M., Lundegaard, C., Lund, O. and Petersen, T. N. (2010). CPHmodels-3.0--remote homology modeling using structure-guided sequence profiles. *Nucleic Acids Res*, **38**, W576-581.

O'Brien, J. B., McCabe, M. S., Athie-Morales, V., McDonald, G. S., Ni Eidhin, D. B. and Kelleher, D. P. (2005). Passive immunisation of hamsters against Clostridium difficile infection using antibodies to surface layer proteins. *FEMS Microbiol Lett*, **246**, 199-205.

O'Connor, J. R., Lyras, D., Farrow, K. A., Adams, V., Powell, D. R., Hinds, J., Cheung, J. K. and Rood, J. I. (2006). Construction and analysis of chromosomal Clostridium difficile mutants. *Mol Microbiol*, **61**, 1335-1351.

O'Hara, A. M. and Shanahan, F. (2006). The gut flora as a forgotten organ. *EMBO Rep*, **7**, 688-693.

O'Horo, J. and Safdar, N. (2009). The role of immunoglobulin for the treatment of Clostridium difficile infection: a systematic review. *Int J Infect Dis*, **13**, 663-667.

Onderdonk, A. B., Lowe, B. R. and Bartlett, J. G. (1979). Effect of environmental stress on Clostridium difficile toxin levels during continuous cultivation. *Appl Environ Microbiol*, **38**, 637-641.

Osserman, E. F. and Lawlor, D. P. (1966). Serum and urinary lysozyme (muramidase) in monocytic and monomyelocytic leukemia. *J Exp Med*, **124**, 921-952.

Otto, H. H. and Schirmeister, T. (1997). Cysteine Proteases and Their Inhibitors. *Chem Rev*, **97**, 133-172.

Pal, K., Kumar, S., Sharma, S., Garg, S. K., Alam, M. S., Xu, H. E., Agrawal, P. and Swaminathan, K. (2010). Crystal structure of full-length Mycobacterium tuberculosis H37Rv glycogen branching enzyme: insights of N-terminal beta-sandwich in substrate specificity and enzymatic activity. *J Biol Chem*, **285**, 20897-20903.

Panessa-Warren, B. J., Tortora, G. T. and Warren, J. B. (2007). High resolution FESEM and TEM reveal bacterial spore attachment. *Microsc Microanal*, **13**, 251-266.

Pantosti, A., Cerquetti, M., Viti, F., Ortisi, G. and Mastrantonio, P. (1989). Immunoblot analysis of serum immunoglobulin G response to surface proteins of Clostridium difficile in patients with antibiotic-associated diarrhea. *J Clin Microbiol*, **27**, 2594-2597.

Parsons, H. K., Vitovski, S. and Sayers, J. R. (2004). Immunoglobulin A1 proteases: a structure-function update. *Biochem Soc Trans*, **32**, 1130-1132.

Pavkov, T., Egelseer, E. M., Tesarz, M., Svergun, D. I., Sleytr, U. B. and Keller, W. (2008). The structure and binding behavior of the bacterial cell surface layer protein SbsC. *Structure*, **16**, 1226-1237.

Pechine, S., Gleizes, A., Janoir, C., Gorges-Kergot, R., Barc, M. C., Delmee, M. and Collignon, A. (2005a). Immunological properties of surface proteins of Clostridium difficile. *J Med Microbiol*, **54**, 193-196.

Pechine, S., Janoir, C., Boureau, H., Gleizes, A., Tsapis, N., Hoys, S., Fattal, E. and Collignon, A. (2007). Diminished intestinal colonization by Clostridium difficile and immune response in mice after mucosal immunization with surface proteins of Clostridium difficile. *Vaccine*, **25**, 3946-3954.

Pechine, S., Janoir, C. and Collignon, A. (2005b). Variability of Clostridium difficile surface proteins and specific serum antibody response in patients with Clostridium difficile-associated disease. *J Clin Microbiol*, **43**, 5018-5025.

Peebles, C. L., Perlman, P. S., Mecklenburg, K. L., Petrillo, M. L., Tabor, J. H., Jarrell, K. A. and Cheng, H. L. (1986). A self-splicing RNA excises an intron lariat. *Cell*, **44**, 213-223.

Peersen, O. B., Ruggles, J. A. and Schultz, S. C. (2002). Dimeric structure of the Oxytricha nova telomere end-binding protein alpha-subunit bound to ssDNA. *Nat Struct Biol*, **9**, 182-187.

Peltier, J., El Meouche, I., Camiade, E., Dupuy, B., Lemee, L. and Pons, J. L. (2010). Autolysins of Clostridium difficile involved in vegitative growth. 3rd International Clostridium difficile symposium, Bled, Slovenia.

Pepin, J., Routhier, S., Gagnon, S. and Brazeau, I. (2006). Management and outcomes of a first recurrence of Clostridium difficile-associated disease in Quebec, Canada. *Clin Infect Dis*, **42**, 758-764.

Pepin, J., Valiquette, L. and Cossette, B. (2005). Mortality attributable to nosocomial Clostridium difficile-associated disease during an epidemic caused by a hypervirulent strain in Quebec. *Cmaj*, **173**, 1037-1042.

Perutka, J., Wang, W., Goerlitz, D. and Lambowitz, A. M. (2004). Use of computer-designed group II introns to disrupt Escherichia coli DExH/D-box protein and DNA helicase genes. *J Mol Biol*, **336**, 421-439.

Petersen, T. N., Kauppinen, S. and Larsen, S. (1997). The crystal structure of rhamnogalacturonase A from Aspergillus aculeatus: a right-handed parallel beta helix. *Structure*, **5**, 533-544.

Potapov, V., Sobolev, V., Edelman, M., Kister, A. and Gelfand, I. (2004). Protein--protein recognition: juxtaposition of domain and interface cores in immunoglobulins and other sandwich-like proteins. *J Mol Biol*, **342**, 665-679.

Potempa, J. and Pike, R. N. (2005). Bacterial peptidases. *Contrib Microbiol*, **12**, 132-180.

Potempa, J., Sroka, A., Imamura, T. and Travis, J. (2003). Gingipains, the major cysteine proteinases and virulence factors of Porphyromonas gingivalis: structure, function and assembly of multidomain protein complexes. *Curr Protein Pept Sci*, **4**, 397-407.

Potempa, M., Potempa, J., Kantyka, T., Nguyen, K. A., Wawrzonek, K., Manandhar, S. P., Popadiak, K., Riesbeck, K., Eick, S. and Blom, A. M. (2009). Interpain A, a cysteine proteinase from Prevotella intermedia, inhibits complement by degrading complement factor C3. *PLoS Pathog*, **5**, e1000316.

Pothoulakis, C., Sullivan, R., Melnick, D. A., Triadafilopoulos, G., Gadenne, A. S., Meshulam, T. and LaMont, J. T. (1988). Clostridium difficile toxin A stimulates intracellular calcium release and chemotactic response in human granulocytes. *J Clin Invest*, **81**, 1741-1745.

Poxton, I. R. and Cartmill, T. D. (1982). Immunochemistry of the cell-surface carbohydrate antigens of Clostridium difficile. *J Gen Microbiol*, **128**, 1365-1370.

Poxton, I. R., McCoubrey, J. and Blair, G. (2001). The pathogenicity of Clostridium difficile. *Clin Microbiol Infect*, **7**, 421-427.

Prieto, C. I., Rodriguez, M. E., Bosch, A., Chirdo, F. G. and Yantorno, O. M. (2003). Whole-bacterial cell enzyme-linked immunosorbent assay for cell-bound Moraxella bovis pili. *Vet Microbiol*, **91**, 157-168.

Pruitt, R. N., Chagot, B., Cover, M., Chazin, W. J., Spiller, B. and Lacy, D. B. (2009). Structure-function analysis of inositol hexakisphosphate-induced autoprocessing in Clostridium difficile toxin A. *J Biol Chem*, **284**, 21934-21940.

Pruitt, R. N., Chambers, M. G., Ng, K. K., Ohi, M. D. and Lacy, D. B. (2010). Structural organization of the functional domains of Clostridium difficile toxins A and B. *Proc Natl Acad Sci U S A*, **107**, 13467-13472.

Purdy, D., O'Keeffe, T. A., Elmore, M., Herbert, M., McLeod, A., Bokori-Brown, M., Ostrowski, A. and Minton, N. P. (2002). Conjugative transfer of clostridial shuttle vectors from Escherichia coli to Clostridium difficile through circumvention of the restriction barrier. *Mol Microbiol*, **46**, 439-452.

Puri, A. W., Lupardus, P. J., Deu, E., Albrow, V. E., Garcia, K. C., Bogyo, M. and Shen, A. (2010). Rational design of inhibitors and activity-based probes targeting Clostridium difficile virulence factor TcdB. *Chem Biol*, **17**, 1201-1211.

Qazi, O., Hitchen, P., Tissot, B., Panico, M., Morris, H. R., Dell, A. and Fairweather, N. (2009). Mass spectrometric analysis of the S-layer proteins from Clostridium difficile demonstrates the absence of glycosylation. *J Mass Spectrom*, **44**, 368-374.

Ramirez, N., Liggins, M. and Abel-Santos, E. (2010). Kinetic evidence for the presence of putative germination receptors in Clostridium difficile spores. *J Bacteriol*, **192**, 4215-4222.

Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G. and Deber, C. M. (2009). Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proc Natl Acad Sci U S A*, **106**, 1760-1765.

Rawlings, N. D. and Barrett, A. J. (1993). Evolutionary families of peptidases. *Biochem J*, **290 (Pt 1)**, 205-218.

Rawlings, N. D., Barrett, A. J. and Bateman, A. (2010). MEROPS: the peptidase database. *Nucleic Acids Res*, **38**, D227-233.

Razaq, N., Sambol, S., Nagaro, K., Zukowski, W., Cheknis, A., Johnson, S. and Gerding, D. N. (2007). Infection of hamsters with historical and epidemic BI types of Clostridium difficile. *J Infect Dis*, **196**, 1813-1819.

Redelings, M. D., Sorvillo, F. and Mascola, L. (2007). Increase in Clostridium difficile-related mortality rates, United States, 1999-2004. *Emerg Infect Dis*, **13**, 1417-1419.

Rediers, H., Rainey, P. B., Vanderleyden, J. and De Mot, R. (2005). Unraveling the secret lives of bacteria: use of in vivo expression technology and differential fluorescence induction promoter traps as tools for exploring niche-specific gene expression. *Microbiol Mol Biol Rev*, **69**, 217-261.

Reynolds, C. B., Emerson, J. E., de la Riva, L., P., F. R. and Fairweather, N. F. (2010). Variation and conservation: The regulation of Clostridium difficile cell wall protein CwpV. 3rd Interational Clostridium difficile symposium, Bled, Slovenia.

Ries, W., Hotzy, C., Schocher, I., Sleytr, U. B. and Sara, M. (1997). Evidence that the N-terminal part of the S-layer protein from Bacillus stearothermophilus PV72/p2 recognizes a secondary cell wall polymer. *J Bacteriol*, **179**, 3892-3898.

Roberts, A. P., Hennequin, C., Elmore, M., Collignon, A., Karjalainen, T., Minton, N. and Mullany, P. (2003). Development of an integrative vector for the expression of antisense RNA in Clostridium difficile. *J Microbiol Methods*, **55**, 617-624.

Rollins, S. M., Peppercorn, A., Young, J. S., Drysdale, M., Baresch, A., Bikowski, M. V., Ashford, D. A., Quinn, C. P., Handfield, M., Hillman, J. D., Lyons, C. R., Koehler, T. M., Calderwood, S. B. and Ryan, E. T. (2008). Application of in vivo induced antigen technology (IVIAT) to Bacillus anthracis. *PLoS One*, **3**, e1824.

Rostand, K. S. and Esko, J. D. (1997). Microbial adherence to and invasion through proteoglycans. *Infect Immun*, **65**, 1-8.

Rothfuss, H., Lara, J. C., Schmid, A. K. and Lidstrom, M. E. (2006). Involvement of the S-layer proteins Hpi and SlpA in the maintenance of cell envelope integrity in Deinococcus radiodurans R1. *Microbiology*, **152**, 2779-2787.

Rudin, A., Olbe, L. and Svennerholm, A. M. (1996). Monoclonal antibodies against fimbrial subunits of colonization factor antigen I (CFA/I) inhibit binding to human enterocytes and protect against enterotoxigenic Escherichia coli expressing heterologous colonization factors. *Microb Pathog*, **21**, 35-45.

Rupnik, M., Dupuy, B., Fairweather, N. F., Gerding, D. N., Johnson, S., Just, I., Lyerly, D. M., Popoff, M. R., Rood, J. I., Sonenshein, A. L., Thelestam, M., Wren, B. W., Wilkins, T. D. and von Eichel-Streiber, C. (2005). Revised nomenclature of Clostridium difficile toxins and associated genes. *J Med Microbiol*, **54**, 113-117.

Rupnik, M., Grabnar, M. and Geric, B. (2003). Binary toxin producing Clostridium difficile strains. *Anaerobe*, **9**, 289-294.

Sabet, M., Lee, S. W., Nauman, R. K., Sims, T. and Um, H. S. (2003). The surface (S-) layer is a virulence factor of Bacteroides forsythus. *Microbiology*, **149**, 3617-3627.

Sakakibara, J., Nagano, K., Murakami, Y., Higuchi, N., Nakamura, H., Shimozato, K. and Yoshimura, F. (2007). Loss of adherence ability to human gingival epithelial cells in S-layer protein-deficient mutants of Tannerella forsythensis. *Microbiology*, **153**, 866-876.

Salcedo, J., Keates, S., Pothoulakis, C., Warny, M., Castagliuolo, I., LaMont, J. T. and Kelly, C. P. (1997). Intravenous immunoglobulin therapy for severe Clostridium difficile colitis. *Gut*, **41**, 366-370.

Salnikova, M. S., Joshi, S. B., Rytting, J. H., Warny, M. and Middaugh, C. R. (2008). Physical characterization of clostridium difficile toxins and toxoids: effect of the formaldehyde crosslinking on thermal stability. *J Pharm Sci*, **97**, 3735-3752.

Sambol, S. P., Tang, J. K., Merrigan, M. M., Johnson, S. and Gerding, D. N. (2001). Infection of hamsters with epidemiologically important strains of Clostridium difficile. *J Infect Dis*, **183**, 1760-1766.

Sanchez-Hurtado, K., Corretge, M., Mutlu, E., McIlhagger, R., Starr, J. M. and Poxton, I. R. (2008). Systemic antibody response to Clostridium difficile in colonized patients with and without symptoms and matched controls. *J Med Microbiol*, **57**, 717-724.

Sanchez-Hurtado, K. and Poxton, I. R. (2008). Enhancement of the cytotoxic activity of Clostridium difficile toxin A by surface-associated antigens. *J Med Microbiol*, **57**, 739-744.

Sanchez-Puelles, J. M., Ronda, C., Garcia, J. L., Garcia, P., Lopez, R. and Garcia, E. (1986). Searching for autolysin functions. Characterization of a pneumococcal mutant deleted in the lytA gene. *Eur J Biochem*, **158**, 289-293.

Sara, M. and Sleytr, U. B. (2000). S-Layer proteins. *J Bacteriol*, **182**, 859-868.

Savariau-Lacomme, M. P., Lebarbier, C., Karjalainen, T., Collignon, A. and Janoir, C. (2003). Transcription and analysis of polymorphism in a cluster of genes encoding surface-associated proteins of Clostridium difficile. *J Bacteriol*, **185**, 4461-4470.

Schaffer, C. and Messner, P. (2005). The structure of secondary cell wall polymers: how Gram-positive bacteria stick their cell walls together. *Microbiology*, **151**, 643-651.

Schell, M. A. (1993). Molecular biology of the LysR family of transcriptional regulators. *Annu Rev Microbiol*, **47**, 597-626.

Scheuring, S., Stahlberg, H., Chami, M., Houssin, C., Rigaud, J. L. and Engel, A. (2002). Charting and unzipping the surface layer of Corynebacterium glutamicum with the atomic force microscope. *Mol Microbiol*, **44**, 675-684.

Schwan, C., Stecher, B., Tzivelekidis, T., van Ham, M., Rohde, M., Hardt, W. D., Wehland, J. and Aktories, K. (2009). Clostridium difficile toxin CDT induces formation of microtubule-based protrusions and increases adherence of bacteria. *PLoS Pathog*, **5**, e1000626.

Scott, M. E., Melton-Celsa, A. R. and O'Brien, A. D. (2003). Mutations in hns reduce the adherence of Shiga toxin-producing E. coli 091:H21 strain B2F1 to human colonic epithelial cells and increase the production of hemolysin. *Microb Pathog*, **34**, 155-159.

Sebaihia, M., Wren, B. W., Mullany, P., Fairweather, N. F., Minton, N., Stabler, R., Thomson, N. R., Roberts, A. P., Cerdeno-Tarraga, A. M., Wang, H., Holden, M. T., Wright, A., Churcher, C., Quail, M. A., Baker, S., Bason, N., Brooks, K., Chillingworth, T., Cronin, A., Davis, P., Dowd, L., Fraser, A., Feltwell, T., Hance, Z., Holroyd, S., Jagels, K., Moule, S., Mungall, K., Price, C., Rabbinowitsch, E., Sharp, S., Simmonds, M., Stevens, K., Unwin, L., Whithead, S., Dupuy, B., Dougan, G., Barrell, B. and Parkhill, J. (2006). The multidrug-resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome. *Nat Genet*, **38**, 779-786.

Seddon, S. V. and Borriello, S. P. (1992). Proteolytic activity of Clostridium difficile. *J Med Microbiol*, **36**, 307-311.

Sekot, G., Posch, G., Messner, P., Matejka, M., Rausch-Fan, X., Andrukhov, O. and Schaffer, C. (2011). Potential of the Tannerella forsythia S-layer to delay the immune response. *J Dent Res*, **90**, 109-114.

Serrano, R. L., Kuhn, A., Hendricks, A., Helms, J. B., Sinning, I. and Groves, M. R. (2004). Structural analysis of the human Golgi-associated plant pathogenesis related protein GAPR-1 implicates dimerization as a regulatory mechanism. *J Mol Biol*, **339**, 173-183.

Sheahan, K. L., Cordero, C. L. and Satchell, K. J. (2007). Autoprocessing of the Vibrio cholerae RTX toxin by the cysteine protease domain. *Embo J*, **26**, 2552-2561.

Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H. and Mann, M. (1996). Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc Natl Acad Sci U S A*, **93**, 14440-14445.

Shi, J. and Casanova, J. E. (2006). Invasion of host cells by Salmonella typhimurium requires focal adhesion kinase and p130Cas. *Mol Biol Cell*, **17**, 4698-4708.

Shida, T., Hattori, H., Ise, F. and Sekiguchi, J. (2001). Mutational analysis of catalytic sites of the cell wall lytic N-acetylmuramoyl-L-alanine amidases CwlC and CwlV. *J Biol Chem*, **276**, 28140-28146.

Shim, J. K., Johnson, S., Samore, M. H., Bliss, D. Z. and Gerding, D. N. (1998). Primary symptomless colonisation by Clostridium difficile and decreased risk of subsequent diarrhoea. *Lancet*, **351**, 633-636.

Shire, S. J., Shahrokh, Z. and Liu, J. (2004). Challenges in the development of high protein concentration formulations. *J Pharm Sci*, **93**, 1390-1402.

Siani, H., Groen, H., Maillard, J. Y. and Baillie, L. (2010). Characterisation of variant C. difficile morphotypes - Microscopic examination and proteome analysis. 3rd Interational Clostridium difficile symposium, Bled, Slovina.

Silva, E. F., Medeiros, M. A., McBride, A. J., Matsunaga, J., Esteves, G. S., Ramos, J. G., Santos, C. S., Croda, J., Homma, A., Dellagostin, O. A., Haake, D. A., Reis, M. G. and Ko, A. I. (2007). The terminal portion of leptospiral immunoglobulin-like protein LigA confers protective immunity against lethal infection in the hamster model of leptospirosis. *Vaccine*, **25**, 6277-6286.

Sleytr, U. B. and Glauert, A. M. (1976). Ultrastructure of the cell walls of two closely related clostridia that possess different regular arrays of surface subunits. *J Bacteriol*, **126**, 869-882.

Sleytr, U. B. and Messner, P. (1983). Crystalline surface layers on bacteria. *Annu Rev Microbiol*, **37**, 311-339.

Smit, E., Oling, F., Demel, R., Martinez, B. and Pouwels, P. H. (2001). The S-layer protein of Lactobacillus acidophilus ATCC 4356: identification and characterisation of domains responsible for S-protein assembly and cell wall binding. *J Mol Biol*, **305**, 245-257.

Soding, J. (2005). Protein homology detection by HMM-HMM comparison. *Bioinformatics*, **21**, 951-960.

Soding, J., Biegert, A. and Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*, **33**, W244-248.

Sorg, J. A. and Sonenshein, A. L. (2010). Inhibiting the Initiation of Clostridium difficile Spore Germination using Analogs of Chenodeoxycholic Acid, a Bile Acid. *J Bacteriol*, **192**, 4983-4990.

Sougioultzis, S., Kyne, L., Drudy, D., Keates, S., Maroo, S., Pothoulakis, C., Giannasca, P. J., Lee, C. K., Warny, M., Monath, T. P. and Kelly, C. P. (2005). Clostridium difficile toxoid vaccine in recurrent C. difficile-associated diarrhea. *Gastroenterology*, **128**, 764-770.

Spigaglia, P., Barbanti, F., Galeotti, C. L., Scarselli, M., Van Broeck, J. and Mastrantonio, P. (2010). Analysis of surface layer (S-layer) proteins of the hypervirulent Clostridium difficile PCR-Ribotype 027 in comparision with those of other ribotypes. 3rd International Clostridium difficile symposium, Bled, Slovenia.

Spigaglia, P., Barbanti, F. and Mastrantonio, P. (2011a). Surface Layer Protein A Variant of Clostridium difficile PCR-Ribotype 027. *Emerg Infect Dis*, **17**, 317-319.

Spigaglia, P., Galeotti, C. L., Barbanti, F., Scarselli, M., Van Broeck, J. and Mastrantonio, P. (2011b). Clostridium difficile PCR-ribotypes 027 and 001 share common immunogenic properties of the low-molecular-weight (LMW) surface layer (S-layer) protein. *J Med Microbiol*.

Stabler, R. A., Gerding, D. N., Songer, J. G., Drudy, D., Brazier, J. S., Trinh, H. T., Witney, A. A., Hinds, J. and Wren, B. W. (2006). Comparative phylogenomics of Clostridium difficile reveals clade specificity and microevolution of hypervirulent strains. *J Bacteriol*, **188**, 7297-7305.

Stabler, R. A., He, M., Dawson, L., Martin, M., Valiente, E., Corton, C., Lawley, T. D., Sebaihia, M., Quail, M. A., Rose, G., Gerding, D. N., Gibert, M., Popoff, M. R., Parkhill, J., Dougan, G. and Wren, B. W. (2009). Comparative genome and phenotypic analysis of Clostridium difficile 027 strains provides insight into the evolution of a hypervirulent bacterium. *Genome Biol*, **10**, R102.

Steinbuch, M. and Audran, R. (1969). The isolation of IgG from mammalian sera with the aid of caprylic acid. *Arch Biochem Biophys*, **134**, 279-284.

Stranger-Jones, Y. K., Bae, T. and Schneewind, O. (2006). Vaccine assembly from surface proteins of Staphylococcus aureus. *Proc Natl Acad Sci U S A*, **103**, 16942-16947.

Strub, M. P., Hoh, F., Sanchez, J. F., Strub, J. M., Bock, A., Aumelas, A. and Dumas, C. (2003). Selenomethionine and selenocysteine double labeling strategy for crystallographic phasing. *Structure*, **11**, 1359-1367.

Sukumar, N., Love, C. F., Conover, M. S., Kock, N. D., Dubey, P. and Deora, R. (2009). Active and passive immunizations with Bordetella colonization factor A protect mice against respiratory challenge with Bordetella bronchiseptica. *Infect Immun*, **77**, 885-895.

Sundriyal, A., Roberts, A. K., Shone, C. C. and Acharya, K. R. (2009). Structural basis for substrate recognition in the enzymatic component of ADP-ribosyltransferase toxin CDTa from Clostridium difficile. *J Biol Chem*, **284**, 28713-28719.

Sunenshine, R. H. and McDonald, L. C. (2006). Clostridium difficile-associated disease: new challenges from an established pathogen. *Cleve Clin J Med*, **73**, 187-197.

Suter-Crazzolara, C. and Unsicker, K. (1995). Improved expression of toxic proteins in E. coli. *Biotechniques*, **19**, 202-204.

Suzuki, N., Yamazaki, Y., Brown, R. L., Fujimoto, Z., Morita, T. and Mizuno, H. (2008). Structures of pseudechetoxin and pseudecin, two snake-venom cysteine-rich secretory proteins that target cyclic nucleotide-gated ion channels: implications for movement of the C-terminal cysteine-rich domain. *Acta Crystallogr D Biol Crystallogr*, **64**, 1034-1042.

Takeoka, A., Takumi, K., Koga, T. and Kawata, T. (1991). Purification and characterization of S layer proteins from Clostridium difficile GAI 0714. *J Gen Microbiol*, **137**, 261-267.

Takumi, K., Endo, Y., Koga, T., Oka, T. and Natori, Y. (1992). In vitro self-assembly of the S layer subunits from Clostridium difficile GAI 0714 into tetragonal arrays. *Tokushima J Exp Med*, **39**, 95-100.

Takumi, K., Takeoka, A. and Kawata, T. (1987). Purification and immunochemical properties of a wall protein antigen from Clostridium difficile ATCC 11011. *Microbiol Immunol*, **31**, 837-849.

Takumi, K., Tetsuro, K., Tatsuzo, O. and Yaeta, E. (1991). Self-assembly, adhesion, and chemical properties of tetragonarly arrayed s-layer proteins of Clostridium. *J. Gen. Appl. Microbiol*, **37**, 455-465.

Tan, K. S., Wee, B. Y. and Song, K. P. (2001). Evidence for holin function of tcdE gene in the pathogenicity of Clostridium difficile. *J Med Microbiol*, **50**, 613-619.

Tanner, H. E., Hardy, K. J. and Hawkey, P. M. (2010). Coexistence of multiple multilocus variable-number tandem-repeat analysis subtypes of Clostridium difficile PCR ribotype 027 strains within fecal specimens. *J Clin Microbiol*, **48**, 985-987.

Tasteyre, A., Barc, M. C., Collignon, A., Boureau, H. and Karjalainen, T. (2001a). Role of FliC and FliD flagellar proteins of Clostridium difficile in adherence and gut colonization. *Infect Immun*, **69**, 7937-7940.

Tasteyre, A., Karjalainen, T., Avesani, V., Delmee, M., Collignon, A., Bourlioux, P. and Barc, M. C. (2001b). Molecular characterization of fliD gene encoding flagellar cap and its expression among Clostridium difficile isolates from different serogroups. *J Clin Microbiol*, **39**, 1178-1183.

Taylor, C. P., Tummala, S., Molrine, D., Davidson, L., Farrell, R. J., Lembo, A., Hibberd, P. L., Lowy, I. and Kelly, C. P. (2008). Open-label, dose escalation phase I study in healthy volunteers to evaluate the safety and pharmacokinetics of a human monoclonal antibody to Clostridium difficile toxin A. *Vaccine*, **26**, 3404-3409.

Teixeira, L. M., Strickland, A., Mark, S. S., Bergkvist, M., Sierra-Sastre, Y. and Batt, C. A. (2010). Entropically driven self-assembly of Lysinibacillus sphaericus S-layer proteins analyzed under various environmental conditions. *Macromol Biosci*, **10**, 147-155.

Tendeng, C. and Bertin, P. N. (2003). H-NS in Gram-negative bacteria: a family of multifaceted proteins. *Trends Microbiol*, **11**, 511-518.

Thompson, S. A. (2002). Campylobacter surface-layers (S-layers) and immune evasion. *Ann Periodontol*, **7**, 43-53.

Todar, K. (2008). "Clostridium: Todar's Online Textbook of Bacteriology." http://www.textbookofbacteriology.net/. 2008.

Toro, N., Jimenez-Zurdo, J. I. and Garcia-Rodriguez, F. M. (2007). Bacterial group II introns: not just splicing. *FEMS Microbiol Rev*, **31**, 342-358.

Torres, J. F., Lyerly, D. M., Hill, J. E. and Monath, T. P. (1995). Evaluation of formalin-inactivated Clostridium difficile vaccines administered by parenteral and mucosal routes of immunization in hamsters. *Infect Immun*, **63**, 4619-4627.

Tsuge, K., Ano, T., Hirai, M., Nakamura, Y. and Shoda, M. (1999). The genes degQ, pps, and lpa-8 (sfp) are responsible for conversion of Bacillus subtilis 168 to plipastatin production. *Antimicrob Agents Chemother*, **43**, 2183-2192.

Tu, Z. C., Gaudreau, C. and Blaser, M. J. (2005). Mechanisms underlying Campylobacter fetus pathogenesis in humans: surface-layer protein variation in relapsing infections. *J Infect Dis*, **191**, 2082-2089.

Tummuru, M. K. and Blaser, M. J. (1992). Characterization of the Campylobacter fetus sapA promoter: evidence that the sapA promoter is deleted in spontaneous mutant strains. *J Bacteriol*, **174**, 5916-5922.

Tung, J. M., Dolovich, L. R. and Lee, C. H. (2009). Prevention of Clostridium difficile infection with Saccharomyces boulardii: a systematic review. *Can J Gastroenterol*, **23**, 817-821.

Turner, M. S. and Helmann, J. D. (2000). Mutations in multidrug efflux homologs, sugar isomerases, and antimicrobial biosynthesis genes differentially elevate activity of the sigma(X) and sigma(W) factors in Bacillus subtilis. *J Bacteriol*, **182**, 5202-5210.

Twine, S. M., Reid, C. W., Aubry, A., McMullin, D. R., Fulton, K. M., Austin, J. and Logan, S. M. (2009). Motility and flagellar glycosylation in Clostridium difficile. *J Bacteriol*, **191**, 7050-7062.

Umemoto, J., Bhavanandan, V. P. and Davidson, E. A. (1977). Purification and properties of an endo-alpha-N-acetyl-D-galactosaminidase from Diplococcus pneumoniae. *J Biol Chem*, **252**, 8609-8614.

Underwood, S., Guan, S., Vijayasubhash, V., Baines, S. D., Graham, L., Lewis, R. J., Wilcox, M. H. and Stephenson, K. (2009). Characterization of the sporulation initiation pathway of Clostridium difficile and its role in toxin production. *J Bacteriol*, **191**, 7296-7305.

Vaishnavi, C. (2009). Established and potential risk factors for Clostridum difficile infection. *Indian J Med Microbiol*, **27**, 289-300.

Vernet, T., Berti, P. J., de Montigny, C., Musil, R., Tessier, D. C., Menard, R., Magny, M. C., Storer, A. C. and Thomas, D. Y. (1995). Processing of the papain precursor. The ionization state of a conserved amino acid motif within the Pro region participates in the regulation of intramolecular processing. *J Biol Chem*, **270**, 10838-10846.

Vernet, T., Khouri, H. E., Laflamme, P., Tessier, D. C., Musil, R., Gour-Salin, B. J., Storer, A. C. and Thomas, D. Y. (1991). Processing of the papain precursor. Purification of the zymogen and characterization of its mechanism of processing. *J Biol Chem*, **266**, 21451-21457.

Viscidi, R., Laughon, B. E., Yolken, R., Bo-Linn, P., Moench, T., Ryder, R. W. and Bartlett, J. G. (1983). Serum antibody response to toxins A and B of Clostridium difficile. *J Infect Dis*, **148**, 93-100.

Viscidi, R., Willey, S. and Bartlett, J. G. (1981). Isolation rates and toxigenic potential of Clostridium difficile isolates from various patient populations. *Gastroenterology*, **81**, 5-9.

Vollmer, W., Joris, B., Charlier, P. and Foster, S. (2008). Bacterial peptidoglycan (murein) hydrolases. *FEMS Microbiol Rev*, **32**, 259-286.

Vonberg, R. P., Kuijper, E. J., Wilcox, M. H., Barbut, F., Tull, P., Gastmeier, P., van den Broek, P. J., Colville, A., Coignard, B., Daha, T., Debast, S., Duerden, B. I., van den Hof, S., van der Kooi, T., Maarleveld, H. J., Nagy, E., Notermans, D. W., O'Driscoll, J., Patel, B., Stone, S. and Wiuff, C. (2008). Infection control measures to limit the spread of Clostridium difficile. *Clin Microbiol Infect*, **14 Suppl 5**, 2-20.

Voth, D. E. and Ballard, J. D. (2005). Clostridium difficile toxins: mechanism of action and role in disease. *Clin Microbiol Rev*, **18**, 247-263.

Wada, N., Nishida, N., Iwaki, S., Ohi, H., Miyawaki, T., Taniguchi, N. and Migita, S. (1980). Neutralizing activity against Clostridium difficile toxin in the supernatants of cultured colostral cells. *Infect Immun*, **29**, 545-550.

Waligora, A. J., Barc, M. C., Bourlioux, P., Collignon, A. and Karjalainen, T. (1999). Clostridium difficile cell attachment is modified by environmental factors. *Appl Environ Microbiol*, **65**, 4234-4238.

Waligora, A. J., Hennequin, C., Mullany, P., Bourlioux, P., Collignon, A. and Karjalainen, T. (2001). Characterization of a cell surface protein of Clostridium difficile with adhesive properties. *Infect Immun*, **69**, 2144-2153.

Wang, G., Maier, S. E., Lo, L. F., Maier, G., Dosi, S. and Maier, R. J. (2010). Peptidoglycan deacetylation in Helicobacter pylori contributes to bacterial survival by mitigating host immune responses. *Infect Immun*.

Wang, L. and Lin, M. (2008). A novel cell wall-anchored peptidoglycan hydrolase (autolysin), IspC, essential for Listeria monocytogenes virulence: genetic and proteomic analysis. *Microbiology*, **154**, 1900-1913.

Warny, M., Vaerman, J. P., Avesani, V. and Delmee, M. (1994). Human antibody response to Clostridium difficile toxin A in relation to clinical course of infection. *Infect Immun*, **62**, 384-389.

Watanabe, K., Hata, Y., Kizaki, H., Katsube, Y. and Suzuki, Y. (1997). The refined crystal structure of Bacillus cereus oligo-1,6-glucosidase at 2.0 A resolution: structural characterization of proline-substitution sites for protein thermostabilization. *J Mol Biol*, **269**, 142-153.

Wei, Y., Guffanti, A. A., Ito, M. and Krulwich, T. A. (2000). Bacillus subtilis YqkI is a novel malic/Na+-lactate antiporter that enhances growth on malate at low protonmotive force. *J Biol Chem*, **275**, 30287-30292.

Weiss, K. (2009). Toxin-binding treatment for Clostridium difficile: a review including reports of studies with tolevamer. *Int J Antimicrob Agents*, **33**, 4-7.

Wexler, H., Mulligan, M. E. and Finegold, S. M. (1984). Polyacrylamide gel electrophoresis patterns produced by Clostridium difficile. *Rev Infect Dis*, **6 Suppl 1**, S229-234.

Wiener, M. C. (2004). A pedestrian guide to membrane protein crystallization. *Methods*, **34**, 364-372.

Wierenga, R. K. (2001). The TIM-barrel fold: a versatile framework for efficient enzymes. *FEBS Lett*, **492**, 193-198.

Wilcox, M. and Minton, J. (2001). Role of antibody response in outcome of antibiotic-associated diarrhoea. *Lancet*, **357**, 158-159.

Wilcox, M. H. (2004). Descriptive study of intravenous immunoglobulin for the treatment of recurrent Clostridium difficile diarrhoea. *J Antimicrob Chemother*, **53**, 882-884.

Williams, D. R., Young, D. I. and Young, M. (1990). Conjugative plasmid transfer from Escherichia coli to Clostridium acetobutylicum. *J Gen Microbiol*, **136**, 819-826.

Wilson, G. G. and Murray, N. E. (1991). Restriction and modification systems. *Annu Rev Genet*, **25**, 585-627.

Wright, A., Drudy, D., Kyne, L., Brown, K. and Fairweather, N. F. (2008). Immunoreactive cell wall proteins of Clostridium difficile identified by human sera. *J Med Microbiol*, **57**, 750-756.

Wright, A., Wait, R., Begum, S., Crossett, B., Nagy, J., Brown, K. and Fairweather, N. (2005). Proteomic analysis of cell surface proteins from Clostridium difficile. *Proteomics*, **5**, 2443-2452.

Yamamoto, Y., Kurata, M., Watabe, S., Murakami, R. and Takahashi, S. Y. (2002). Novel cysteine proteinase inhibitors homologous to the proregions of cysteine proteinases. *Curr Protein Pept Sci*, **3**, 231-238.

Yao, J., Zhong, J. and Lambowitz, A. M. (2005). Gene targeting using randomly inserted group II introns (targetrons) recovered from an Escherichia coli gene disruption library. *Nucleic Acids Res*, **33**, 3351-3362.

Yao, Q., Cui, J., Zhu, Y., Wang, G., Hu, L., Long, C., Cao, R., Liu, X., Huang, N., Chen, S., Liu, L. and Shao, F. (2009). A bacterial type III effector family uses the papain-like hydrolytic activity to arrest the host cell cycle. *Proc Natl Acad Sci U S A*, **106**, 3716-3721.

Yeats, C., Bentley, S. and Bateman, A. (2003). New knowledge from old: in silico discovery of novel protein domains in Streptomyces coelicolor. *BMC Microbiol*, **3**, 3.

Yeats, C., Rawlings, N. D. and Bateman, A. (2004). The PepSY domain: a regulator of peptidase activity in the microbial environment? *Trends Biochem Sci*, **29**, 169-172.

Yoneda, M., Hirofuji, T., Motooka, N., Nozoe, K., Shigenaga, K., Anan, H., Miura, M., Kabashima, H., Matsumoto, A. and Maeda, K. (2003). Humoral immune responses to S-layer-like proteins of Bacteroides forsythus. *Clin Diagn Lab Immunol*, **10**, 383-387.

Yoshida, M., Claypool, S. M., Wagner, J. S., Mizoguchi, E., Mizoguchi, A., Roopenian, D. C., Lencer, W. I. and Blumberg, R. S. (2004). Human neonatal Fc receptor mediates transport of IgG into luminal secretions for delivery of antigens to mucosal dendritic cells. *Immunity*, **20**, 769-783.

Zhang, D., Li, N., Lok, S. M., Zhang, L. H. and Swaminathan, K. (2003). Isomaltulose synthase (PalI) of Klebsiella sp. LX3. Crystal structure and implication of mechanism. *J Biol Chem*, **278**, 35428-35434.

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, **9**, 40.

Zhong, J., Karberg, M. and Lambowitz, A. M. (2003). Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. *Nucleic Acids Res*, **31**, 1656-1664.